Adaptive estimation of the dynamics of a discrete time stochastic volatility model

Comte, F.; Lacour, C.; Rozenholc, Y.

Accepted Manuscript

Adaptive estimation of the dynamics of a discrete time stochastic volatility model

F. Comte, C. Lacour, Y. Rozenholc

PII: S0304-4076(09)00150-X
DOI: 10.1016/j.jeconom.2009.07.001
Reference: ECONOM 3216

To appear in: *Journal of Econometrics*

Received date: 31 January 2008
Revised date: 23 June 2009
Accepted date: 2 July 2009

Please cite this article as: Comte, F., Lacour, C., Rozenholc, Y., Adaptive estimation of the dynamics of a discrete time stochastic volatility model. *Journal of Econometrics* (2009), doi:10.1016/j.jeconom.2009.07.001

ADAPTIVE ESTIMATION OF THE DYNAMICS OF A DISCRETE TIME STOCHASTIC VOLATILITY MODEL

F. COMTE[1], C. LACOUR[2], AND Y. ROZENHOLC[1]

Abstract. This paper is concerned with the discrete time stochastic volatility model $Y_i = \exp(X_i/2)\eta_i$, $X_{i+1} = b(X_i) + \sigma(X_i)\xi_{i+1}$, where only (Y_i) is observed. The model is re-written as a particular hidden model: $Z_i = X_i + \varepsilon_i$, $X_{i+1} = b(X_i) + \sigma(X_i)\xi_{i+1}$, where (ξ_i) and (ε_i) are independent sequences of i.i.d. noise. Moreover, the sequences (X_i) and (ε_i) are independent and the distribution of ε is known. Then, our aim is to estimate the functions b and σ^2 when only observations $Z_1, \ldots, Z_n$ are available. We propose to estimate bf and $(b^2+\sigma^2)f$ and study the integrated mean square error of projection estimators of these functions on automatically selected projection spaces. By ratio strategy, estimators of b and σ^2 are then deduced. The mean square risk of the resulting estimators are studied and their rates are discussed. Lastly, simulation experiments are provided: constants in the penalty functions defining the estimators are calibrated and the quality of the estimators is checked on several examples.

J.E.L. Classification number: C13-C14-C22.

Keywords. Adaptive Estimation; Autoregression; Deconvolution; Heteroscedastic; Hidden Markov Model; Nonparametric Projection Estimator.

1. Introduction

In this paper, we consider the following model:

$$\begin{cases} Y_i = \exp(X_i/2)\eta_i, \\ X_{i+1} = b(X_i) + \sigma(X_i)\xi_{i+1}, \end{cases} \tag{1}$$

where (η_i) and (ξ_i) are two independent sequences of independent and identically distributed (i.i.d.) random variables (noise processes). Only $Y_1, \ldots, Y_n$ are observed, while the process of interest is the unobserved volatility $V_i = \exp(X_i/2)$, and in particular the functions driving its dynamics, $b(.)$ and $\sigma(.)$. We will describe an estimation method leading to nonparametric estimates of these functions.

For identifiability of the model, the density of η must be known (e.g. $\mathcal{N}(0,1)$). This model is often called a discrete time stochastic volatility process. Examples of such representation for economic or financial processes can be found in Shephard (2005) or Shephard (2006).

[1] Université Paris Descartes, MAP5, UMR CNRS 8145.
fabienne.comte@parisdescartes.fr, yves.rozenholc@parisdescartes.fr.
[2] Laboratoire de Probabilités et Statistique, Université Paris Sud-Orsay.
claire.lacour@math.u-psud.fr.

1.1. **Comparison with continuous time stochastic volatility models.** Let us emphasize that our model is written directly in discrete time with fixed (to 1) step of observation.

This makes its structure different from stochastic volatility models when first considered in continuous time and in a second step discretely sampled. It would be conceivable to study a system

$$\left\{ \begin{array}{lcl} d\log(S_t) & = & \sqrt{U_t}dW_t \\ dU_t & = & m(U_t)dt + \varpi(U_t)dB_t, \end{array} \right. \tag{2}$$

where (W_t, B_t) is a two dimensional (standard) Brownian motion and U is a positive diffusion process. In that case, define $Y_i = \log(S_{i+1}/S_i)$. Then, Y_i has the same distribution as $V_i\eta_i$ where $V_i = (\int_i^{i+1} U_s ds)^{1/2}$ and η_i are i.i.d. $\mathcal{N}(0,1)$ random variables. Consequently, the first equation of (2) can lead exactly to the first equation of (1) with specific Gaussian distribution for η. The tools for estimating the common density of the V_i's are thus common to both models. But the second equations of models (1) and (2), even if they represent the same idea of a time dynamics, do not coincide.

Indeed, many statistical tools have been developed for the estimation of μ and ϖ, see Renò (2008), Bandi and Renò (2008), Comte *et al.* (2009) in the nonparametric setting, or Gloter (2008) in the parametric context. But all these authors consider the high frequency data context, a set of assumptions which are often meaningful, but specific to observations with small step $\Delta = \Delta_n$ that tends to zero when n grows to infinity, and long interval of time: $T_n = n\Delta_n$ tends to infinity with n. Most strategies in this context require a huge sample of observations, which may not be available. Moreover, the discrete time approximations of the variables crucially exploit the small step assumption. Therefore, this does not correspond to the discrete time setting with fixed step of observation. Very few papers use another context to provide estimators of the continuous time model. Following ideas developed in Hansen and Scheinkman (1987) and later in Hansen *et al.* (1998), Gobet *et al.* (2004) propose a strategy for low frequency data (i.e. fixed sample step Δ of observation). They obtain nonparametric estimators in this context, but their method is specific to the underlying continuous time model. The methodology is similar to the operator eigenvalues approximation studied in Carrasco *et al.* (2007) but can not be applied to our model.

Consequently, Model (1), and in particular the autoregressive equation in (1), is not an approximation of the continuous time model (2) (it would require a small sample step for an approximation to hold) and does not share any underlying continuous time structure.

1.2. **Logarithmic transformation of (1) and related models.** When the model is described in discrete time only, estimation strategies have been proposed for the estimation of the stationary density of (X_i) only, see van Es *et al.* (2005), Comte *et al.* (2008). In this case, the model is rewritten as

$$\left\{ \begin{array}{l} Z_i = X_i + \varepsilon_i \\ X_{i+1} = b(X_i) + \sigma(X_i)\xi_{i+1} \end{array} \right. \tag{3}$$

where $\varepsilon_i = \ln(\eta_i^2) - \mathbb{E}(\ln(\eta_i^2))$ and $Z_i = \ln(Y_i^2) - \mathbb{E}(\ln(\eta_i^2))$. In our setting, $\mathbb{E}(\ln(\eta_i^2))$ is known. Clearly, the logarithmic transformation of Y_i^2 implies that the sign of Y_i can not be recovered. When we are interested in volatility dynamics, the sign of the innovation in Y does not matter since we have the additional assumption of independent volatility and price innovation, i.e. (η) and (ξ) are independent. All our results require this independence assumption, and thus, this context does not authorize the so-called leverage effect (see also Remark 3.1).

It also happens that in some contexts, model (3) is considered directly. Such a non linear autoregressive model observed with additive noise is also called an errors-in-variables model and is commonly considered in economic and biological applications (see Schenmach (2007), Hong and Tamer (2003) or Benett and Wakefield (2001)). Statistical methods used to study these models are mainly related to parametric or semiparametric specifications of the model, see Chanda (1995), Comte and Taupin (2001).

Lastly, Model (3) belongs to the general class of hidden Markov models (HMM). These models constitute a very famous class of discrete time processes with applications in various areas (see Cappé, Moulines and Ryden (2005)). Here our model is simpler in the sense that our noise is additive, but in standard HMMs it is assumed that the joint density of (X_i, Z_i) has a parametric form.

1.3. **Statistical bibliography for model (3).** To our knowledge, the question of estimating b and σ^2 in Model (3) on the basis of observations $Z_1, \ldots, Z_n$ has not been studied yet.

Only the following regressive model $Z_i = X_i + \varepsilon_i$, $Y_i = b(X_i) + \xi_i$, in which (Y_i) and (Z_i) for $i = 1, \ldots, n+1$ are observed, has received attention. In that case, two processes are observed, all sequences (X_i), (ξ_i), (ε_i) can be supposed independent identically distributed (i.i.d.) and independent from each other, and (Y_i) is homoscedastic ($\sigma(x) \equiv 1$). In this context, Fan and Truong (1990), and Comte and Taupin (2007) study the problem of the estimation of b. See also Fan *et al.* (1991), Fan and Masry (1992), Ioannides and Alevizos (1997), Koo and Lee (1998). Most authors propose estimators of b based on the ratio of two estimators, namely an estimator of bf divided by an estimator of f, where f denotes the common density of the (i.i.d. in their context) (X_i).

Several papers develop estimation methods for f, see Fan (1991), Pensky and Vidakovic (1999), Comte *et al.* (2006), Carrasco *et al.* (2007), and the optimality of the rates are studied in Fan (1991), Butucea (2004) and Butucea and Tsybakov (2007).

1.4. **Statistical strategy.** The quotient strategy is also adopted in our more general setting. More precisely, we assume that the process (X_i) is stationary, with stationary density denoted by f, and we estimate b (resp. $b^2 + \sigma^2$) as a ratio of an estimator of bf (resp. $(b^2 + \sigma^2)f$) divided by an estimator of f. We estimate f with the adaptive estimator proposed by Comte *et al.* (2006). To estimate the numerators, we adopt the same type of strategy as for the estimation of f, namely: first, we study a projection estimator, then we propose a model selection criterion to choose the best projection space as possible, in term of mean integrated risk bound.

In the setting of (3), regarding the identifiability of the model, it must be assumed that the distribution of ε, f_ε (or equivalently of η) is fully known. For instance, the process η is often modeled as a standard Gaussian i.i.d. sequence, and then ε_i has the distribution of $\ln(\mathcal{N}(0,1)^2) + \ln(2) + C$ where C is the Euler constant. Van Es *et al.* (2005) specifically study this case in terms of density estimation, however more general distributions can also be considered (see Comte *et al.* (2006, 2007)).

In this respect, this allows to consider general classes of noise density f_ε and also various classes of regularities for the functions to estimate $(bf,\ (b^2 + \sigma^2)f,\ f)$. Then, when all functions belong to fixed (but user-unknown) regularity spaces, our risk bounds provide rates of convergence of the estimators.

We want here to emphasize the following point. After the pioneering work of Fan (1991), people mainly remembered that the estimation of a twice differentiable density measured with an additive gaussian noise ε had a deplorable logarithmic rate. This is true, but often, the functions to recover are much more regular than only twice differentiable. If the signal has the same regularity as the noise, then the rate can become polynomial again (see Example 3.1 below). This was found by Carrasco *et al.* (2007) with their specific method and also by Pensky and Vidakovic (1999), Comte *et al* (2006), or Butucea and Tsybakov (2008) with wavelet, projection or kernel estimators respectively.

The plan of the paper is the following: we first give the notations, the assumptions and describe projection spaces in Section 2. Next, Section 3 explains the estimation strategy for b and gives bounds of the integrated mean square risk of the estimators. Section 4 develops the same study for the estimation of σ^2. Simulation experiments are conducted in Section 5 in order to illustrate the method. Lastly, proofs are gathered in Sections 6-7-8 and an appendix (section 9) describes auxiliary tools.

2. General setting and assumptions

2.1. **The principle.** Let us assume that the sequence (X_i) is stationary and let us denote by f the common density of the X_i's. The principle of the estimation methods relies in all cases on a "Nadaraya-Watson-strategy" in the sense that b or $b^2+\sigma^2$ are estimated as ratio of an estimator of $\ell = bf$ (respectively $\vartheta = (b^2+\sigma^2)f$) and an estimator of f.
In all cases, we use the adaptive estimator of f described in Comte *et al.* (2006) or Comte it et al. (2008) which study independent and β-mixing contexts.

2.2. **Notations and Assumptions.** Subsequently we denote by u^* the Fourier transform of the function u defined by $u^*(t) = \int e^{itx}u(x)dx$, and by $\|u\|$, $\|u\|_\infty$, $\|u\|_{\infty,K}$, $< u, v >$, $u * v$ the quantities

$$\|u\|^2 = \int u^2(x)dx,\ \|u\|_\infty = \sup_{x\in\mathbb{R}}|u(x)|,\ \|u\|_{\infty,K} = \sup_{x\in K}|u(x)|,$$

$$< u, v >= \int u(x)\overline{v}(x)dx \text{ with } z\overline{z} = |z|^2 \text{ and } u * v(x) = \int u(t)\bar{v}(x-t)dt.$$

Moreover, we recall that for any integrable and square-integrable functions u, u_1, u_2,

$$(u^*)^*(x) = 2\pi u(-x) \text{ and } \langle u_1, u_2\rangle = (2\pi)^{-1}\langle u_1^*, u_2^*\rangle. \tag{4}$$

We consider the autoregressive model (3). The assumptions are the following:

- **A1** (a) The ε_i's are i.i.d. centered ($\mathbb{E}(\varepsilon_1) = 0$) random variables with finite variance, $\mathbb{E}(\varepsilon_1^2) = s_\varepsilon^2$. The density of ε_1, f_ε, belongs to $\mathbb{L}_2(\mathbb{R})$, and for all $x \in \mathbb{R}$, $f_\varepsilon^*(x) \neq 0$.
  (b) The ξ_i's are i.i.d. centered with unit variance ($\mathbb{E}(\xi_1^2) = 1$).
- **A2** The X_i's are stationary and absolutely regular.
- **A3** The sequences $(\xi_i)_{i\in\mathbb{N}}$ and $(\varepsilon_i)_{i\in\mathbb{N}}$ are independent.

It follows from **A3** and the data generating process of the X_i that the sequences $(X_i)_{i\in\mathbb{N}}$ and $(\varepsilon_i)_{i\in\mathbb{N}}$ are independent.

The Z_i's are observed but the X_i's are not, the stationary density f of the X_i's is unknown but the density f_ε of the ε_i's is known.

Standard assumptions on b, σ and the ξ_i's ensure that the sequence $(X_i)_{i\in\mathbb{Z}}$ is stationary with stationary density denoted by f. This sequence is also absolutely regular, with β-mixing coefficients denoted by $\beta(k)$, see Doukhan (1994) or Comte and Rozenholc (2002) for precise sets of conditions. We shall consider that the mixing is at least arithmetical with rate θ, i.e. that there exists $\theta > 0$ such that

$$\forall k \in \mathbb{N},\ \beta(k) \le (1+k)^{-(1+\theta)}, \tag{5}$$

or, more often, geometrical, i.e. $\exists \theta > 0, \forall k \in \mathbb{N}, \beta(k) \le e^{-\theta k}$. The definition of the β-mixing coefficients and related properties are recalled in Section 9.

As we develop a $\mathbb{L}^2$-strategy, we need the target functions to be square-integrable.

A4 The function to estimate ($\ell = bf$, $\vartheta = (b^2+\sigma^2)f$, or f) is square-integrable.

In the following, we also assume that f_ε is such that

A5 For all t in $\mathbb{R}$, $A_0(t^2+1)^{-\gamma/2}\exp\{-\mu|t|^\delta\} \le |f_\varepsilon^*(t)| \le A_0'(t^2+1)^{-\gamma/2}\exp\{-\mu|t|^\delta\}$, with $\gamma > 1/2$ if $\delta = 0$.

Under Assumption **A5**, when $\delta = 0$, the errors are usually called "ordinary smooth" errors, and "super smooth" errors when $\delta > 0, \mu > 0$. The standard examples are the following : Gaussian or Cauchy distributions are super smooth of order $(\gamma = 0, \mu = 1/2, \delta = 2)$ and $(\gamma = 0, \mu = 1, \delta = 1)$ respectively, and the Laplace (symmetric exponential) distribution is ordinary smooth ($\delta = 0$) of order $\gamma = 2$. When $\varepsilon = \ln(\eta^2) - \mathbb{E}(\ln(\eta^2))$ with $\eta \sim \mathcal{N}(0,1)$ as in van Es *et al.* (2005), then $\mathbb{E}(\ln(\eta^2)) = -\ln(2) - C$, where C is the Euler Constant, and $\mathrm{Var}(\ln(\eta^2)) = \pi^2/2$ and ε is super-smooth with $\gamma = 0$, $\mu = \pi/2$ and $\delta = 1$:

$$f^*_{\ln(\eta_1^2)}(x) = \frac{2^{ix}}{\sqrt{\pi}}\Gamma(1+ix),\ \ |f^*_{\ln(\eta_1^2)}(x)| \sim_{x\to+\infty} \sqrt{2/e}e^{-\pi|x|/2}.$$

2.3. **The projection spaces.** As projection estimators are used in all cases, we hereby provide a description of the projection spaces. Let us define

$$\varphi(x) = \frac{\sin(\pi x)}{\pi x} \ \text{ and } \ \varphi_{m,j}(x) = \sqrt{m}\varphi(mx-j),$$

where m can be replaced by 2^m. The key point is that $\varphi^*(x) = \mathbf{I}_{[-\pi,\pi]}(x)$. It is well known (see Meyer (1990), p.22) that $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$ is an orthonormal basis of the space of square integrable functions having a Fourier transform with compact support included into $[-\pi m, \pi m]$. Such a space is denoted by S_m.

$$S_m \ = \ \mathrm{Span}\{\varphi_{m,j},\ j\in\mathbb{Z}\} \ = \{f \in \mathbb{L}^2(\mathbb{R}), \mathrm{supp}(f^*) \subset [-m\pi, m\pi]\}.$$

Moreover, $(S_m)_{m\in\mathcal{M}_n}$, with $\mathcal{M}_n = \{1,\ldots,m_n\}$, denotes the collection of linear spaces.

In practice, we should consider the truncated spaces $S_m^{(n)} = \mathrm{Span}\{\varphi_{m,j},\ \ j\in\mathbb{Z},\ \ |j| \le K_n\}$, where K_n is an integer depending on n, and the associated estimators under the additional assumption: $\int x^2\psi^2(x)dx < A_\psi < \infty$, where $\psi = bf, (b^2+\sigma^2)f$ or f is the function to estimate. This is done in Comte *et al.* (2006) and does not change the main part of the study. For the sake of simplicity, we write the sums over $\mathbb{Z}$ in the theoretical part of the present study.

3. Estimation of b

3.1. The steps of the estimation.

3.1.1. *First step: the estimators of $\ell = bf$.* The orthogonal projection of $\ell = bf$ on S_m, ℓ_m, is given by

$$\ell_m = \sum_{j\in\mathbb{Z}} a_{m,j}(\ell)\varphi_{m,j} \text{ with } a_{m,j}(\ell) = \int_{\mathbb{R}} \varphi_{m,j}(x)\ell(x)dx = \langle \varphi_{m,j}, \ell\rangle. \tag{6}$$

For t belonging to a space S_m of the collection $(S_m)_{m\in\mathcal{M}_n}$, let

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^n (\|t\|^2 - 2Z_{i+1}u_t^*(Z_i)), \quad u_t(x) = \frac{1}{2\pi}\frac{t^*(-x)}{f_\varepsilon^*(x)}.$$

The following sequence of equalities, relying on the Fourier equalities (4), explains the choice of the contrast γ_n:

$$\begin{aligned}
\mathbb{E}(Z_2u_t^*(Z_1)) &= \mathbb{E}(b(X_1)u_t^*(Z_1)) = \langle u_t^* * f_\varepsilon(-.), bf\rangle = \frac{1}{2\pi}\langle \frac{t^*}{f_\varepsilon^*(-.)} f_\varepsilon^*(-.), (bf)^*\rangle \\
&= \frac{1}{2\pi}\langle t^*, (bf)^*\rangle = \langle t, bf\rangle = \mathbb{E}(b(X_1)t(X_1)) = \int t(x)b(x)f(x)dx \\
&= \langle t, \ell\rangle, \qquad (7)
\end{aligned}$$

using that

$$\mathbb{E}(\sigma(X_1)\xi_2 u_t^*(X_1+\varepsilon_1)) = \mathbb{E}(\xi_2)\mathbb{E}(\sigma(X_1)u_t^*(X_1+\varepsilon_1)) = 0. \tag{8}$$

Therefore, we find that

$$\mathbb{E}(\gamma_n(t)) = \|t\|^2 - 2\langle \ell, t\rangle = \|t-\ell\|^2 - \|\ell\|^2$$

is minimal when $t = \ell$. Thus, we define

$$\hat{\ell}_m = \arg\min_{t\in S_m} \gamma_n(t) \tag{9}$$

As γ_n is minimized over S_m only, $\hat{\ell}_m$ is in fact an unbiased estimator of ℓ_m, which in turn is expected to be near of ℓ. Indeed, the estimator can also be written

$$\hat{\ell}_m = \sum_{j\in\mathbb{Z}} \hat{a}_{m,j}(\ell)\varphi_{m,j}, \text{ with } \hat{a}_{m,j}(\ell) = \frac{1}{n}\sum_{i=1}^n Z_{i+1}u^*_{\varphi_{m,j}}(Z_i), \tag{10}$$

an clearly, (7) implies that $\mathbb{E}(\hat{a}_{m,j}(\ell)) = a_{m,j}(\ell)$. Now, the decomposition of the Mean Integrated Squared Error (MISE) will show that m plays here the role of a bandwidth parameter. Thus, we have to explain how to select an adequate value of m. To this end, we define $\hat{\ell}_{\hat{m}}$, by setting

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n} \left\{\gamma_n(\hat{\ell}_m) + \mathrm{pen}(m)\right\},$$

where the penalty function is given by $\mathrm{pen}(m) = \kappa \mathbb{E}(Z_2^2)\Psi(m)$ where

$$\Psi(m) = \begin{cases} \dfrac{\Delta(m)}{n} & \text{if } 0 \leq \delta < 1/3 \\ \dfrac{m^{[(3\delta-1)/2]\wedge\delta}\Delta(m)}{n} & \text{if } \delta \geq 1/3, \end{cases} \quad \text{and } \Delta(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{dx}{|f_\varepsilon^*(x)|^2}, \tag{11}$$

where $x \wedge y := \inf(x, y)$.

The idea behind this criterion is that $\gamma_n(\hat{\ell}_m)$ estimates the squared bias part of the MISE, and $\mathrm{pen}(m)$ has the order of the variance term. A kind of cross-validation method is thus performed here.

In practice $\mathbb{E}(Z_2^2)$ is unknown and is replaced by its empirical version, $(1/n)\sum_{i=1}^n Z_i^2$. Then, the resulting penalty function, $\widehat{\mathrm{pen}}$, becomes random. We note that $\gamma_n(\hat{\ell}_m) = -\sum_{j\in\mathbb{Z}}[\hat{a}_{m,j}(\ell)]^2$, which explains (13) below.

Remark 3.1. *We can see here why we can not omit the independence assumption between (ε_i) and (ξ_i), and consider the so-called leverage effect. Indeed, assume that we have $\xi_i = \rho\varepsilon_i + \sqrt{1-\rho^2}\varepsilon_i^{\perp}$, for an i.i.d. centered noise $(\varepsilon_i^{\perp})$, independent of (ε_i). Then, because of the dynamics of the autoregressive X_i sequence, if the noises (ε_i) and (ξ_i) are not independent, then the sequences (X_i) and (ε_i) are not independent either. Then, as for a given i, X_i and ε_i are no longer independent in model (3), we can not estimate f (the basic convolution link is lost). To keep X_i and ε_i independent for a given i (and then f can still be estimated, see Comte et al. (2007,2008)), we have to write the dynamics of X_i as follows: $X_{i+1} = b(X_i) + \sigma(X_i)\xi_i$. But then, Equality (7) which justifies the definition of the contrast $\gamma_n(t)$ used for the estimation of ℓ, is no longer true. Indeed, instead of (8), we have*

$$\mathbb{E}(\sigma(X_1)\xi_1 u_t^*(X_1+\varepsilon_1)) = \rho\mathbb{E}(\sigma(X_1)\varepsilon_1 u_t^*(X_1+\varepsilon_1)),$$

and this last term is not necessarily zero. As a conclusion, the independence assumption between (ε_i) and (ξ_i) is necessary.

3.1.2. *Second step: the estimators of f.* The second stage of the estimation procedure is to estimate f. In fact, Comte *et al.* (2006) explain how to estimate f in an adaptive way and the mixing context is studied in Comte *et al.* (2008). The estimator of f on S_m is defined by

$$\hat{f}_m = \sum_{j\in\mathbb{Z}} \hat{a}_{m,j}(f)\varphi_{m,j} \quad \text{with} \quad \hat{a}_{m,j}(f) = \frac{1}{n}\sum_{i=1}^n u^*_{\varphi_{m,j}}(Z_i). \tag{12}$$

Then we define $\hat{f}_{\ddot{m}}$,

$$\ddot{m} = \arg\min_{m\in\mathcal{M}_n}\left\{-\sum_{j\in\mathbb{Z}}[\hat{a}_{m,j}(f)]^2 + \ddot{\mathrm{pen}}(m)\right\}, \tag{13}$$

where the penalty function is given by $\ddot{\mathrm{pen}}(m) = \ddot{\kappa}\Psi(m)$ with $\Psi(m)$ given by (11). For the properties of $\hat{f}_{\ddot{m}}$ we refer to Comte *et al.* (2006). Up to the multiplicative constants, the control of the mean square risk of the estimator is the same as the one obtained for ℓ here.

3.1.3. *Last step: the estimator of b.* We estimate b on a compact set B only and the following additional assumption is required:

A6 (a) $\forall x \in B,\ \ f_0 \leq f(x) \leq f_1$ for two positive constants f_0 and f_1.
(b) b is bounded on B.

Then we can define:

$$\tilde{b} = \hat{b}_{\hat{m},\ddot{m}} = \frac{\hat{\ell}_{\hat{m}}}{\hat{f}_{\ddot{m}}} \text{ if } \|\hat{\ell}_{\hat{m}}/\hat{f}_{\ddot{m}}\| \leq a_n,\ \ \tilde{b} = \hat{b}_{\hat{m},\ddot{m}} = 0 \text{ else,} \tag{14}$$

where a_n is a sequence to be specified later.

3.2. Risk bound for $\hat{\ell}_m$ and $\hat{\ell}_{\hat{m}}$.

3.2.1. *Risk bound for $\hat{\ell}_m$.* We define the following empirical centered process

$$\nu_n(t) = \frac{1}{n}\sum_{k=1}^{n}(Z_{k+1}u_t^*(Z_k) - \langle t, \ell\rangle),$$

and with (6) and (10), we note that the following equalities hold

$$\begin{aligned}\|\ell - \hat{\ell}_m\|^2 &= \|\ell - \ell_m\|^2 + \|\ell_m - \hat{\ell}_m\|^2 = \|\ell - \ell_m\|^2 + \sum_{j\in\mathbb{Z}}(a_{m,j}(\ell) - \hat{a}_{m,j}(\ell))^2 \\ &= \|\ell - \ell_m\|^2 + \sum_{j\in\mathbb{Z}}\nu_n^2(\varphi_{m,j}).\end{aligned}$$

Therefore

$$\mathbb{E}\|\ell - \hat{\ell}_m\|^2 \leq \|\ell - \ell_m\|^2 + \sum_{j\in\mathbb{Z}}\mathrm{Var}[\nu_n(\varphi_{m,j})].$$

The term $\|\ell - \ell_m\|^2$ is a deterministic integrated squared bias term. The part corresponding to $\sum_{j\in\mathbb{Z}}\mathrm{Var}[\nu_n(\varphi_{m,j})]$ is the variance term, which has to be bounded. Then, the risk bound on the estimate $\hat{\ell}_m$ is as follows:

Proposition 3.1. *Consider the estimator $\hat{\ell}_m$ of ℓ defined by (9) where $\ell = bf$ with b and f as in Model (3). Then under Assumptions* **A1-A4***, if $\mathbb{E}(b^4(X_1)) < +\infty$ and $\theta > 1$ for arithmetical mixing (see (5)), we have*

$$\mathbb{E}(\|\ell - \hat{\ell}_m\|^2) \leq \|\ell - \ell_m\|^2 + 2\mathbb{E}(Z_2^2)\frac{\Delta(m)}{n} + 8K\frac{m}{n}. \tag{15}$$

Note that the last term m/n in (15) is always smaller that $\Delta(m)/n$ and might have been omitted, up to a less precise constant before $\Delta(m)/n$.

3.2.2. *Rate of the estimator.* When f_ε^* satisfies **A5**, then the order of the variance term is bounded by:

$$C\Delta(m)/n \leq C' m^{2\gamma+1-\delta}\exp(2\mu(\pi m)^\delta)/n.$$

For the squared bias term, we have $\|\ell - \ell_m\|^2 = (2\pi)^{-1}\int_{|x|\geq\pi m}|f^*(x)|^2dx$. To evaluate this term, regularity conditions must be considered for ℓ. We shall assume that ℓ belongs to the space:

$$\mathcal{S}_{s,a,r}(A) = \{u\ :\ \int_{-\infty}^{+\infty}|u^*(x)|^2(x^2+1)^s\exp\{2a|x|^r\}dx \leq A\}, \tag{16}$$

	$\delta = 0$ f_ε ordinary smooth	$\delta > 0$ f_ε supersmooth
$r = 0$ ℓ Sobolev(s)	$\pi\check{m} = O(n^{1/(2s+2\gamma+1)})$ rate $= O(n^{-2s/(2s+2\gamma+1)})$	$\pi\check{m} = [\ln(n)/(2\mu+1)]^{1/\delta}$ rate $= O((\ln(n))^{-2s/\delta})$
$r > 0$ ℓ $\mathcal{C}^\infty$	$\pi\check{m} = [\ln(n)/2a]^{1/r}$ rate $= O\left(\dfrac{\ln(n)^{(2\gamma+1)/r}}{n}\right)$	$\check{m}$ solution of $\check{m}^{2s+2\gamma+1-r}\exp\{2\mu(\pi\check{m})^\delta + 2a\pi^r\check{m}^r\}$ $= O(n)$

Table 1. Choice of $\check{m}$ and corresponding rates under **A1-A5** and if ℓ belongs to $\mathcal{S}_{s,a,r}(A)$ defined by (16).

for nonnegative constants s, a, r and $A > 0$. When $r = 0$, this corresponds to Sobolev spaces of order s, which are classically considered. But we can also think that ℓ may be as regular as f_ε can be. When $r > 0$, $a > 0$, this corresponds to analytic functions, which are called "super-smooth" functions.
When ℓ belongs to a space $\mathcal{S}_{s,a,r}(A)$ defined by (16), then the order of the squared bias is

$$\|\ell - \ell_m\|^2 = (2\pi)^{-1}\int_{|x|\geq \pi m} |f^*(x)|^2 dx \leq C m^{-2s}\exp(-2a(\pi m)^r).$$

In this setting, we obtain from (15) the risk bound terms:

$$C m^{-2s}\exp(-2a(\pi m)^r) + C' m^{2\gamma+1-\delta}\exp(2\mu(\pi m)^\delta)/n.$$

Example 3.1. Consider the example of a Gaussian η, $\eta \sim \mathcal{N}(0,1)$, i.e. $\varepsilon = \ln(\eta^2) - \mathbb{E}(\ln(\eta^2))$ is super-smooth with $\gamma = 0$, $\mu = \pi/2$ and $\delta = 1$. It is true that if ℓ is in a Sobolev i.e. belongs to $\mathcal{S}_{s,a,r}(A)$ with $r = a = 0$, then the best attainable rate is of order $[\log(n)]^{-2s}$. But if ℓ is as regular as f_ε and belongs to $\mathcal{S}_{s,a,r}(A)$ with $s = 0$, $r = 1$, $a > 0$, then the optimal choice is $\pi\check{m} = \log(n)/(\pi + 2a)$ and the resulting rate is polynomial, of order $n^{-2a/(\pi+2a)}$. This is summarized in Proposition 3.2 and Table 1.

Proposition 3.2. *Assume that the Assumptions of Proposition 3.1 hold. In addition, assume that ℓ belongs to a space $\mathcal{S}_{s,a,r}(A)$ defined by (16) and that Assumption* **A5** *is fulfilled. Then the estimate $\hat{\ell}_{\check{m}}$ with $\check{m}$ as in Table 1, has the rates given in Table 1 in terms of its mean square integrated risk $\mathbb{E}(\|\hat{\ell}_m - \ell\|^2)$.*

When $r > 0, \delta > 0$ the value of $\check{m}$ is not explicitly given. It is obtained as the solution of the equation

$$\check{m}^{2s+2\gamma+1-r}\exp\{2\mu(\pi\check{m})^\delta + 2a\pi^r\check{m}^r\} = O(n).$$

For explicit general formulae of the rates in these cases, we refer to Lacour (2006).

We just recall the particular case $\delta = r$, for positive other parameters: this gives the (almost) polynomial rate $n^{-\mu/(a+\mu)}(\log(n))^{-\tau}$ with $\tau = [2\mu\delta + (\delta - 2\gamma - 1)a]/[(a+\mu)\delta]$.

3.3. **Risk bound of $\hat{\ell}_{\hat{m}}$.** These rates enhance the interest of building an estimator for which the choice of the relevant model m is automatically performed. This is done with $\hat{\ell}_{\hat{m}}$, and we can prove the following result:

Theorem 3.1. *Assume that Assumptions* **A1-A4** *hold, that* $\mathbb{E}(b^8(X_1))$*,* $\mathbb{E}(\sigma^8(X_1))$ *and* $\mathbb{E}(\xi_1^8)$ *are finite and that* $\mathbb{E}(\varepsilon_1^6) < +\infty$*. Assume moreover that the process* X *is geometrically* β*-mixing, (or arithmetically* β*-mixing with* $\theta > 14$*) and that the collection* $\mathcal{M}_n$ *is such that for all* $m \in \mathcal{M}_n$*,* $\text{pen}(m) \leq 1$*, then*

$$\mathbb{E}(\|\hat{\ell}_{\hat{m}} - \ell\|^2) \leq C \inf_{m \in \mathcal{M}_n} \left(\|\ell - \ell_m\|^2 + \text{pen}(m)\right) + \frac{C'}{n}.$$

The proof of Theorem 3.1 is sketched in Section 6.

Theorem 3.1 shows that the estimator automatically selects the optimal m when $\delta \leq 1/3$: indeed, in that case, the penalty has exactly the same order as the variance (namely $\Delta(m)/n$). When $\delta > 1/3$, a compromise is still performed, but the penalty is slightly greater than the variance. In an asymptotic setting, this implies a loss in the rate of convergence of the estimator, but this loss can be shown to be negligible with respect to the rates. For discussions on this point, see Comte *et al.* (2006).

3.4. **Risk bounds for $\tilde{b}$.** Comte *et al.* (2008) prove that $\hat{f}_{\ddot{m}}$ satisfies the same inequality as $\hat{\ell}_{\hat{m}}$. We recall their result.

Theorem 3.2. *Assume that Assumptions* **A1-A4** *hold. Assume that the process* X *is geometrically* β*-mixing, (or arithmetically* β*-mixing with* $\theta > 3$ *in (5)) and that the collection* $\mathcal{M}_n$ *is such that for all* $m \in \mathcal{M}_n$*,* $\text{p\ddot{e}n}(m) \leq 1$*, then*

$$\mathbb{E}(\|\hat{f}_{\ddot{m}} - f\|^2) \leq C \inf_{m \in \mathcal{M}_n} \left(\|f - f_m\|^2 + \text{p\ddot{e}n}(m)\right) + \frac{C'}{n}$$

where f_m *denotes the orthogonal projection of* f *on* S_m*.*

Then it is common (see e.g. Lacour (2008) or Comte and Taupin (2007)) to obtain a bound on the MISE computed on a compact set, denoted here by B. In the following result, we denote by $\|h\|_B^2 := \int_B h^2(x)dx$, for any square integrable function h on B.

Theorem 3.3. *Assume that the assumptions of Theorems 3.1 and 3.2 and* **A5**-**A6** *hold, that* f *belongs to a space* $\mathcal{S}_{s,a,r}(A)$ *with* $s > 1/2$ *if* $r = 0$ *and* ℓ *to a space* $\mathcal{S}_{s',a',r'}(A')$*, that* $\ln(\ln(n)) \leq m_n \leq (n/\ln(n))^{1/(2\gamma+1)}$ *for* $\hat{f}_{\ddot{m}}$*. Then, for* n *great enough, we have*

$$\mathbb{E}(\|\tilde{b} - b\|_B^2) \leq C_1 \mathbb{E}(\|\hat{f}_{\ddot{m}} - f\|^2) + C_2 \mathbb{E}(\|\hat{\ell}_{\hat{m}} - \ell\|^2) + \frac{C_3}{n}$$

where $a_n = n^\omega$ *with* $\omega > 1/2$ *and* C_1, C_2, C_3 *are constants.*

The MISE bound in Theorem 3.3 means that the rate for estimating b will be the worst between the rates of estimation of f and bf, which are optimal or near-optimal.

The proof of this result is omitted. The reader is referred to Lacour (2008) or Comte and Taupin (2007).

4. Estimation of σ^2

4.1. **Steps of the estimation.** We now aim at estimating σ^2 and we also follow the strategy described in Section 2.1.

First step. We set $\vartheta = (b^2+\sigma^2)f$ and we first estimate ϑ. To this end, we consider the following contrast:

$$\breve{\gamma}_n(t) = \|t\|^2 - \frac{2}{n}\sum_{k=1}^{n}\left(Z_{k+1}^2 - s_\varepsilon^2\right)u_t^*(Z_k). \tag{17}$$

As f_ε is assumed to be known, then so is the variance s_ε^2. Suggestions on how to estimate this quantity can be drawn from Butucea and Matias (2004), but this has a cost and they obtain logarithmic rates of convergence for both the variance estimator and the induced plugged in estimator of the density. Then we define

$$\hat{\vartheta}_m = \arg\min_{t\in S_m}\breve{\gamma}_n(t) \text{ and } \breve{m} = \arg\min_{m\in\mathcal{M}_n}\breve{\gamma}_n(\hat{\vartheta}_n) + \breve{\text{pen}}(m) \tag{18}$$

where $\breve{\text{pen}}(m)$ is a penalty function given by: $\breve{\text{pen}}(m) = \breve{\kappa}\mathbb{E}((Z_2^2 - s_\varepsilon^2)^2)\Psi(m)$, with $\Psi(m)$ given by (11). Again, the expectation $\mathbb{E}[(Z_2^2 - s_\varepsilon^2)^2]$ is replaced by its empirical version in practice, namely $(1/n)\sum_{i=1}^{n}(Z_i^2 - s_\varepsilon^2)^2$.

Second step. As previously, we use as an estimator of f, the estimator $\hat{f}_{\ddot{m}}$ defined by (12)-(13). Its risk is controlled by Theorem 3.2.

Third step. We obtain, by defining, similarly to (14),

$$\widetilde{b^2+\sigma^2} = \frac{\hat{\vartheta}_{\breve{m}}}{\hat{f}_{\ddot{m}}} \text{ if } \|\hat{\vartheta}_{\breve{m}}/\hat{f}_{\ddot{m}}\| \le \breve{a}_n \quad \text{and} \quad 0 \text{ otherwise,}$$

and $\breve{a}_n$ is a sequence to be specified in the same way as a_n for the estimation of b. Clearly, $\widetilde{b^2+\sigma^2}$ is an estimator of $b^2+\sigma^2$. For the study of steps 2 and 3, see Section 3.4.

Fourth step. The estimator of σ^2 must be built by setting

$$\tilde{\sigma}^2 = \widetilde{b^2+\sigma^2} - (\tilde{b})^2.$$

Clearly, as $\|\tilde{\sigma}^2 - \sigma^2\|^2 \le 2\|\widetilde{b^2+\sigma^2} - (b^2+\sigma^2)\|^2 + 2\|b+\tilde{b}\|^2\|b-\tilde{b}\|^2$, the risk of the final estimator is the sum of the risks of the estimators of $b^2+\sigma^2$ and b, provided that b is bounded (by M_b) and $\tilde{b}$ is bounded (by $2M_b$) with probability near of one. The latter step is studied from an empirical point of view only.

4.2. **Risk bounds for $\hat{\vartheta}_m$ and $\hat{\vartheta}_{\breve{m}}$.** It is not difficult to check that $\mathbb{E}(\breve{\gamma}_n(t)) = \|t\|^2 - 2\langle\vartheta, t\rangle$ which justifies the choice of $\breve{\gamma}_n$ given in (17). We can also easily obtain the decomposition $\breve{\gamma}_n(t) - \breve{\gamma}_n(s) = \|t-\vartheta\|^2 - \|s-\vartheta\|^2 - 2\breve{\nu}_n(t-s)$ where

$$\breve{\nu}_n(t) = \frac{1}{n}\sum_{k=1}^{n}\left[(Z_{k+1}^2 - s_\varepsilon^2)u_t^*(Z_k) - \langle t, \vartheta\rangle\right].$$

As for b previously, we can write that

$$\|\hat{\vartheta}_m - \vartheta\|^2 \quad = \quad \|\vartheta_m - \vartheta\|^2 + \sum_{j\in\mathbb{Z}}\breve{\nu}_n^2(\varphi_{m,j}).$$

With the same tools as for the study of ℓ, using a relevant decomposition of the empirical process $\breve{\nu}_n$, we prove (see Section 7) that:

Proposition 4.1. *Consider the estimator $\hat{\vartheta}_m$ of ϑ defined by (18) where $\vartheta = (b^2+\sigma^2)f$ with b, σ and f as in Model (3). Then under Assumptions* **A1-A4***, and if $\xi_2, \varepsilon_1, b^2(X_1)$ and $\sigma^2(X_1)$ admit moments of order 4, then*

$$\mathbb{E}(\|\vartheta - \hat{\vartheta}_m\|_2^2) \leq \|\vartheta - \vartheta_m\|^2 + 4\mathbb{E}[(Z_2^2 - s_\varepsilon^2)^2]\frac{\Delta(m)}{n} + \breve{K}\frac{m}{n}$$

where $\breve{K} = 16\sqrt{2\sum_{k\geq 0}(k+1)\beta(k)\mathbb{E}((b^2(X_1)+\sigma^2(X_1))^4)}$ *if* $\sum_k k\beta(k) < +\infty$.

The empirical processes involved in the decomposition of $\breve{\nu}_n$ are of the same type as the processes studied for the estimation of ℓ. Therefore, we give the risk bound for $\hat{\vartheta}_{\breve{m}}$ but we omit the proof.

Theorem 4.1. *Assume that Assumptions* **A1-A4** *hold, that $\mathbb{E}(b^p(X_1))$, $\mathbb{E}(\sigma^p(X_1))$ and $\mathbb{E}(\xi_1^p)$ are finite for a $p \geq 16$ and that $\mathbb{E}(\varepsilon_1^{12}) < +\infty$. Assume that the process X is geometrically β-mixing and that the collection $\mathcal{M}_n$ is such that for all $m \in \mathcal{M}_n$, $\breve{\mathrm{pen}}(m) \leq 1$, then*

$$\mathbb{E}(\|\hat{\vartheta}_{\breve{m}} - \vartheta\|^2) \leq C \inf_{m\in\mathcal{M}_n}\left(\|\vartheta - \vartheta_m\|^2 + \breve{\mathrm{pen}}(m)\right) + \frac{C'}{n}.$$

5. Simulation results

5.1. **The models.** We start from the first equation of the multiplicative model, $Y_i = \exp(X_i/2)\eta_i,$, $i = 1,\dots,n$. Then we compute:

$$Z_i = \ln(Y_i^2) - \mathbb{E}(\ln(\eta_1^2)), \;\; i = 1,\dots,n \tag{19}$$

and the Z_i's follow the equation $Z_i = X_i + \varepsilon_i$, where ε_i is centered with variance s_ε^2. We have $\varepsilon_i = \ln(\eta_i^2) - \mathbb{E}(\ln(\eta_1^2))$.

We consider three types of noises: the case ε_i Laplace with variance $s_{\varepsilon,L}^2$, the case ε_i Gaussian $\mathcal{N}(0, \sigma_{\varepsilon,G}^2)$, and the case called hereafter "log$\chi^2(1)$" where η_i in $\mathcal{N}(0,1)$; in that case, $\mathbb{E}(\ln(\eta_1^2)) = -C - \ln(2)$ where C is the Euler constant, and $s_{\varepsilon,\chi}^2 = \pi^2/2$. In practice, we generate directly the ε_i's. To compute the Z_i's, we need also to generate the X_i's.

We want to experiment different models, thus we have to choose functions b, σ. In all cases, the ξ_i's are i.i.d. $\mathcal{N}(0,1)$ random variables. Then we consider the following functions, defining three models

$$\left\{\begin{array}{ll} b_1^0(x) = 0.25x & \sigma_1^0(x) = 3, \\ b_2^0(x) = 0.25\sin(2\pi x + \pi/3) & \sigma_2^0(x) = (\sqrt{2}/7)(0.31 + 0.7\exp(-5x^2)), \\ b_3^0(x) = -0.25(x + 2*\exp(-16x^2)) & \sigma_3^0(x) = 0.2 + 0.4\exp(-2x^2). \end{array}\right. \tag{20}$$

The first model is a simple linear model, the two other ones are arbitrary chosen to lead to different types of curves.

We compute recursively, up to initial conditions, a sample $U_1,\dots,U_n$ with $U_{i+1} = b^0(U_i) + \sigma^0(U_i)\xi_{i+1}$, for (b^0, σ^0) taken as one of the (b_j^0, σ_j^0), $j = 1,2,3$. Note that we drop out the first 100 observations to have a stationary sample.

At last, we take $(X_i, b, \sigma) = (U_i, b^0, \sigma^0)$ when the ε_i's are Laplace or Gaussian, and $X_i = U_i/\tilde{s}_\varepsilon$, $b(x) = b^0(\tilde{s}_\varepsilon x)/\tilde{s}_\varepsilon$, $\sigma(x) = \sigma^0(\tilde{s}_\varepsilon x)/\tilde{s}_\varepsilon$ when the ε_i's are log $\chi^2(1)$. The factor $\tilde{s}_\varepsilon$ is a scaling factor computed to have the same signal to noise ratio (i.e. the ratio $\mathrm{Var}(X)/\mathrm{Var}(\varepsilon)$) for the three different noises ε. Therefore we take $s_{\varepsilon,L} = s_{\varepsilon,G} = 1$ and thus $\tilde{s}_\varepsilon = \sqrt{2}/\pi$ for Model 1, and $s_{\varepsilon,L} = s_{\varepsilon,G} = 0.05$ and thus $\tilde{s}_\varepsilon = 0.05\sqrt{2}/\pi$ for Models 2 and 3.

5.2. **Description of the estimation algorithm.** Let us describe the way the procedure works. Consider data $(Y_i)_{1\le i\le n}$ for a given noise η in Model (1). Then $\mathbb{E}(\ln(\eta_1^2))$ is known and we compute Z_i as given by (19). For instance if $\eta_1 \sim \mathcal{N}(0,1)$, we take

$$Z_i = \ln(Y_i^2) + C + \ln(2)$$

where C is the Euler constant.

The first step is to estimate ℓ. Given the Z_i, we compute

$$\hat{a}_{m,j}(\ell) = \frac{1}{n}\sum_{i=1}^n Z_{i+1} u^*_{\varphi_{m,j}}(Z_i) = \frac{1}{2\pi n}\sum_{i=1}^n Z_{i+1}\int e^{-iuZ_i}\frac{\varphi_{m,j}(u)}{f_\varepsilon^*(u)}du$$

for $j = -N, \dots, N$, and N great enough. Note that the $a_{m,j}(\ell)$ are real. Indeed, let $\hat{\varphi}_Z(x) = (1/n)\sum_{k=1}^n Z_{k+1}e^{ixZ_k}$. Then with an elementary change of variable, we obtain

$$\hat{a}_{m,j}(\ell) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\hat{\varphi}_Z(-mu)e^{iju}}{f_\varepsilon^*(mu)}\sqrt{m}du = \frac{1}{\pi}\int_0^{\pi} Re\left(\frac{\hat{\varphi}_Z(-mu)e^{iju}}{f_\varepsilon^*(mu)}\right)\sqrt{m}du,$$

where $Re(z)$ denotes the real part of the complex number z. Then m is selected by minimizing over a grid of possible m's the quantity:

$$\gamma_n(\hat{\ell}_m) + \mathrm{pen}(m) = -\sum_j[\hat{a}_{m,j}(\ell)]^2 + \mathrm{pen}(m)$$

with

$$\mathrm{pen}(m) = \frac{1}{2n}\left(\frac{1}{n}\sum_{i=1}^n Z_i^2\right)\left(1 + \frac{\ln^{2.5}(\pi m)}{(\pi m)(1+s_\varepsilon^2)} + C(m,s_\varepsilon)\right)\Delta(m), \quad \Delta(m) = 2\int_0^{\pi m}\frac{du}{|f_\varepsilon^*(u)|^2}$$

where

$$f_\varepsilon^*(x) = \frac{1}{1+\frac{1}{2}s_{\varepsilon,L}^2x^2}\text{(Laplace)}, \quad f_\varepsilon^*(x) = \exp(-\frac{1}{2}s_{\varepsilon,G}^2x^2)\text{ (Gaussian)} \tag{21}$$

or, for the log-$\chi^2(1)$:

$$f_\varepsilon^*(x) = \frac{1}{\sqrt{\pi}}2^{i(x+C+\ln(2))}\Gamma(1+i(x+C+\ln(2))) \tag{22}$$

and $C(m,s_\varepsilon) = 0$ if ε_1 is Laplace,

$$C(m,s_\varepsilon) = \frac{2}{3}(\pi m)^2 s_{\varepsilon,G}^2 \text{ if } \varepsilon_1 \text{ is Gaussian}, \quad C(m,s_\varepsilon) = \pi m s_{\varepsilon,\chi} \text{ if } \varepsilon_1 \sim \ln(\chi^2(1)).$$

Let us mention that the Fourier transforms f_ε^* above correspond to the densities

- $f_\varepsilon(x) = \exp(-\sqrt{2}|x|/s_{\varepsilon,L})/(\sqrt{2}s_{\varepsilon,L})$ for the centered Laplace with variance $s_{\varepsilon,L}^2$,
- $f_\varepsilon(x) = \exp(-x^2/(2s_{\varepsilon,G}^2))/(\sqrt{2\pi}s_{\varepsilon,G})$ for the $\mathcal{N}(0,s_{\varepsilon,G}^2)$,
- $f_\varepsilon(x) = K(x-\ln(2)-C)$ where C denotes the Euler constant and $K(x) = (1/\sqrt{2\pi})\exp((x-e^x)/2)$, for the centered $\log(\chi^2(1))$ distribution with variance $s_{\varepsilon,\chi} = \pi/\sqrt{2}$.

	Laplace ε		$\log\chi^2(1)$ ε		Gaussian ε	
$n =$	1000	5000	1000	5000	1000	5000
b_1	0.0831	0.0215	0.036	0.00523	0.0781	0.0194
$b_1^2+\sigma_1^2$	9.24	1.22	1.2	0.0603	8.95	1.2
σ_1^2	7.65	1.05	0.428	0.0744	8.01	1.32
b_2	2.73e-4	6.52e-05	2.8e-4	6.12e-05	2.6e-4	6.04e-05
$b_2^2+\sigma_2^2$	1.03e-4	1.89e-05	1.56e-4	2.47e-05	1.21e-4	2e-05
σ_2^2	6.75e-05	1.29e-05	9.76e-05	1.74e-05	7.93e-05	1.36e-05
b_3	0.00383	9.45e-4	0.00459	0.00103	0.00447	8.49e-4
$b_3^2+\sigma_3^2$	0.00406	0.00114	0.00447	0.00142	0.00416	0.00114
σ_3^2	0.00296	8.39e-4	0.00324	0.00106	0.00296	8.96e-4

FIGURE 1. ASE for the estimation of b_j^0 and $(\sigma_j^0)^2$, $j = 1, 2, 3$ given in (20), for 100 replications of the estimation procedure.

The penalties are calibrated so that we recover standard orders m/n when $s_\varepsilon \to 0$. Indeed, setting $s_\varepsilon \to 0$ amounts to set the convolution noise to zero and is a limit case where X_i would be observed.

Moreover, the penalties correspond to the variance order $\Delta(m)/n$, with no loss for Laplace errors, with a loss of order m for the $\log\chi^2(1)$ case (where $\delta = 1$) and a loss of order m^2 in the Gaussian case ($\delta = 2$), see formula (11). The common term $\ln^{2.5}(\pi m)$ is an additional adjustment for small m's, which has no asymptotic weight compared to the other terms.

For the estimation of f, the same procedure applies with

- the function $\hat\varphi_Z$ replaced by $\hat f_Z^*(x) = (1/n)\sum_{k=1}^n e^{ixZ_k}$,
- the factor $(1/n)\sum_{i=}^n Z_i^2$ in the penalty replaced by 1.

For the estimation of ϑ, the same procedure applies with

- the function $\hat\varphi_Z$ replaced by

$$\hat\varphi_{Z^2}(x) = \frac{1}{n}\sum_{k=1}^n (Z_{k+1}^2 - s_\varepsilon^2)e^{ixZ_k},$$

- the factor $(1/n)\sum_{i=}^n Z_i^2$ in the penalty replaced by $(1/n)\sum_{i=1}^n (Z_i^2 - s_\varepsilon^2)^2$.

In the three cases, we select the three values of m among 125 values ranging from $10\ln(n)/(\pi n)$ to 10 and of course, they are not the same in general.

5.3. **Simulation results.** Let us present the results of our simulation experiments.

First, we performed a Monte Carlo study which is reported in Table 1. For each simulation, we compute the average squared error (ASE) at 101 grid points of our adaptive estimator and average these ASEs over 100 replications. It must be noticed that we give here the results for the functions b_j^0, σ_j^0, so that the ASE's can be compared in function of the distribution of the noise ε. We can see that, as already noticed in the density deconvolution setting, there is little difference between Laplace, $\log\chi^2(1)$ and Gaussian ε_i's, in spite of the difference between the theoretical rates. Moreover, it appears clearly that increasing the sample size leads to noticeable improvements of the results.

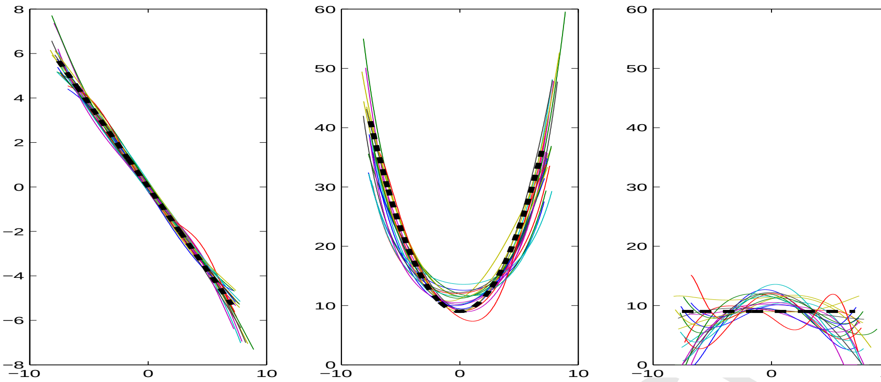

FIGURE 2. True curves: b_1 (left), $b_1^2+\sigma_1^2$ (center), σ_1^2 (right) in dotted-bold and 20 estimated curves, $n = 1000$, ε Laplace.

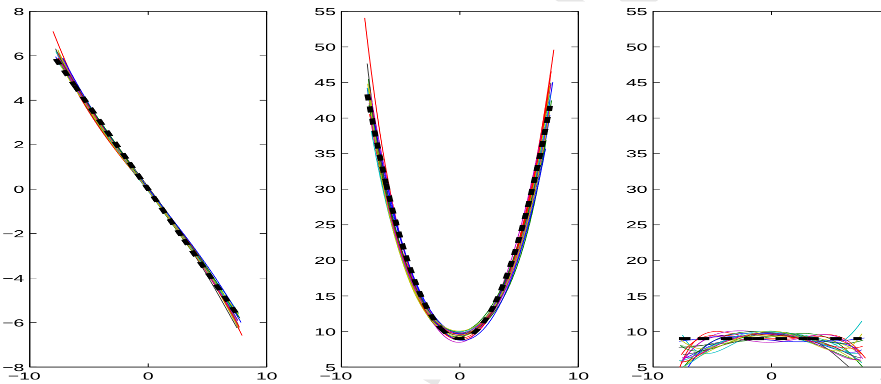

FIGURE 3. True curves: b_1 (left), $b_1^2+\sigma_1^2$ (center), σ_1^2 (right) in dotted-bold and 20 estimated curves, $n = 5000$, ε $\log\chi^2(1)$.

Figures 2 to 6 illustrate our results in the three cases of couples (b_j, σ_j^2) given in (20) with rescaling for $\log\chi^2(1)$ errors ε. The true curves are bold dotted and are given together with 20 estimated curves. This set of twenty estimated curves shows, in each case, that the confidence band around the true function has very small width. Moreover, in all cases, and for all the noise distributions, the estimation are very good. With the financial setting in mind, we took samples with large sizes $n = 1000$ or $n = 5000$ which are clearly good sizes for nonparametric estimation.

We can see that b and $b^2 + \sigma^2$ are well estimated by the ratio strategy. The extraction of σ^2 sometimes suffers from scale problems (if σ^2 is much smaller that b^2 or if both are very small). In particular, we plot σ^2 with the same vertical scale as $b^2 + \sigma^2$ to take this into account.

We can therefore conclude that the method we propose is very precise to estimate b and $b^2 + \sigma^2$, and can give an interesting idea of the shape of σ^2.

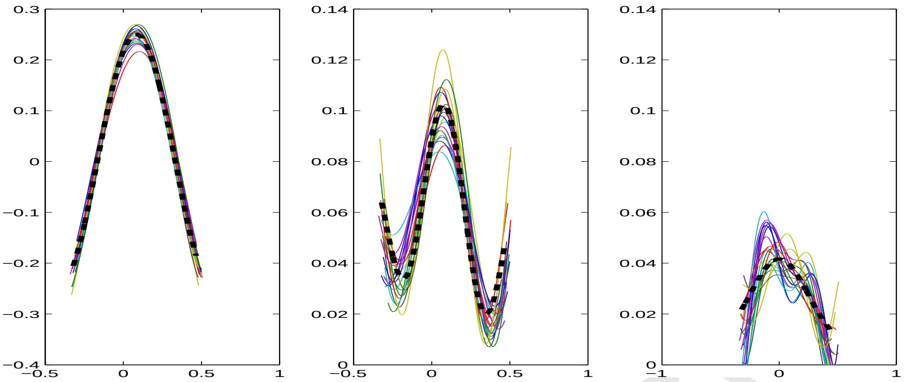

FIGURE 4. True curves: b_2 (left), $b_2^2+\sigma_2^2$ (center), σ_2^2 (right) in dotted-bold and 20 estimated curves, $n = 1000$, ε Laplace.

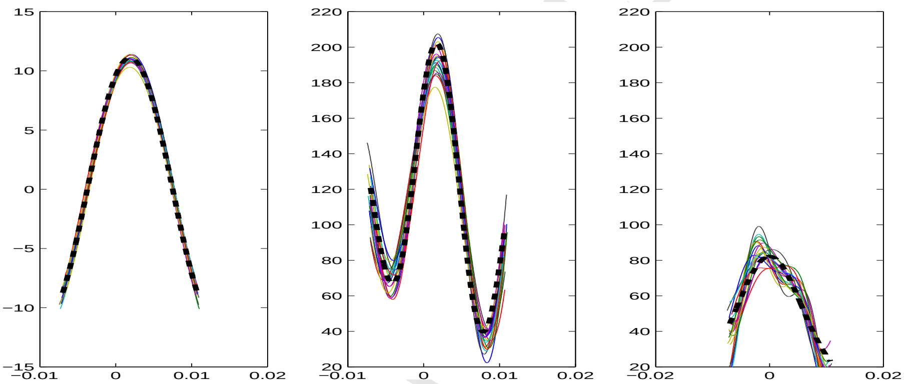

FIGURE 5. True curves: b_2 (left), $b_2^2+\sigma_2^2$ (center), σ_2^2 (right) in dotted-bold and 20 estimated curves, $n = 5000$, $\log\chi^2(1)$-case.

6. PROOFS

6.1. **Proof of Proposition 3.1.** We use the following decomposition: $\nu_n(t) = \nu_n^{(1)}(t) + \nu_n^{(2)}(t) + \nu_n^{(3)}(t)$ with

$$\nu_n^{(1)}(t) = \frac{1}{n}\sum_{k=1}^{n} \varepsilon_{k+1} u_t^*(Z_k), \quad \nu_n^{(2)}(t) = \frac{1}{n}\sum_{k=1}^{n} \xi_{k+1}\sigma(X_k) u_t^*(Z_k), \tag{23}$$

$$\nu_n^{(3)}(t) = \frac{1}{n}\sum_{k=1}^{n} (b(X_k) u_t^*(Z_k) - \langle t, \ell\rangle). \tag{24}$$

Here the terms $\nu_n^{(1)}$ and $\nu_n^{(2)}$ can be kept together and benefit from the uncorrelatedness of the variables involved in the sums. The term $\nu_n^{(3)}$ involves dependent variables. Then

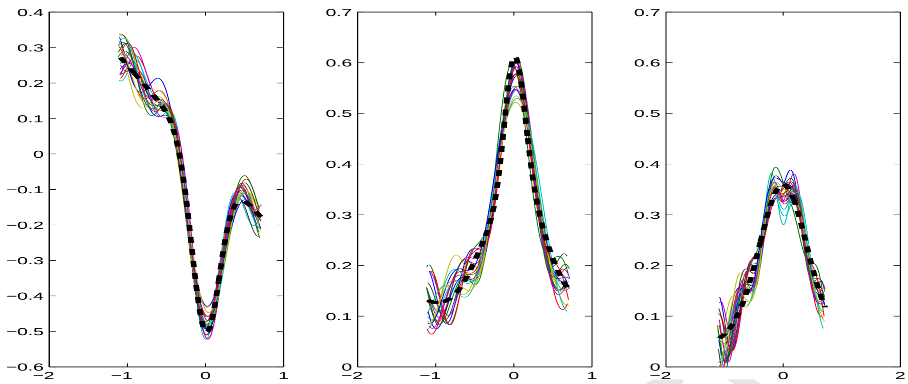

FIGURE 6. True curves: b_3 (left), $b_3^2+\sigma_3^2$ (center), σ_3^2 (right) in dotted-bold and 20 estimated curves, $n = 5000$, Gaussian ε.

we find

$$\mathrm{Var}[\nu_n(\varphi_{m,j})] \le 2\mathrm{Var}\left[\nu_n^{(1)}(\varphi_{m,j}) + \nu_n^{(2)}(\varphi_{m,j})\right] + 2\mathrm{Var}\left[\nu_n^{(3)}(\varphi_{m,j})\right].$$

The first variance involves uncorrelated and centered terms and leads to

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_{i+1} + \sigma(X_i)\xi_{i+1})u^*_{\varphi_{m,j}}(Z_i)\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}[(s_\varepsilon^2 + \sigma^2(X_i))|u^*_{\varphi_{m,j}}(Z_i)|^2]$$

so that

$$\sum_{j\in\mathbb{Z}}\mathrm{Var}\left[\nu_n^{(1)}(\varphi_{m,j}) + \nu_n^{(2)}(\varphi_{m,j})\right] = \frac{1}{n}\sum_{j\in\mathbb{Z}}\mathbb{E}[(s_\varepsilon^2+\sigma^2(X_1))|u^*_{\varphi_{m,j}}(Z_1)|^2] = \frac{(s_\varepsilon^2 + \mathbb{E}(\sigma^2(X_1)))\Delta(m)}{n}.$$

We use here the following useful property of our basis (resulting from a Parseval's formula):

$$\forall x \in \mathbb{R}, \sum_j |u^*_{\varphi_{m,j}}(x)|^2 = \Delta(m),$$

where $\Delta(m)$ is defined by (11) and the $u^*_{\varphi_{m,j}}(x)$ are just rewritten as Fourier coefficients.

For the second term, we use the standard tools specific to the β-mixing context (namely Viennet's (1997) covariance Inequality) and we can easily prove the following Lemma:

Lemma 6.1. *Under Assumptions* **A1-A3**, *and if* $\mathbb{E}(b^4(X_1)) < \infty$ *and* $\sum_k k\beta(k) < +\infty$, *then*

$$\sum_{j\in\mathbb{Z}}\mathrm{Var}\left(\nu_n^{(3)}(\varphi_{m,j})\right) \le \mathbb{E}(b^2(X_1))\frac{\Delta(m)}{n} + \frac{4Km}{n},$$

where $K = \sqrt{2\sum_{k\ge 0}(k+1)\beta(k)\mathbb{E}(b^4(X_1))}$.

The result of Proposition 3.1 follows. □

Proof of Lemma 6.1.

$$\mathrm{Var}\left(\nu_n^{(3)}(\varphi_{m,j})\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}\left(b(X_k)u^*_{\varphi_{m,j}}(Z_k)\right) + \frac{1}{n^2}\sum_{1\le k\ne l\le n}\mathrm{cov}(b(X_k)u^*_{\varphi_{m,j}}(Z_k), b(X_l)u^*_{\varphi_{m,j}}(Z_l)).$$

Then

$$\mathbb{E}(b(X_k)u^*_{\varphi_{m,j}}(Z_k)) = \frac{1}{2\pi}\int \mathbb{E}(b(X_1)e^{ixX_1})\varphi^*_{m,j}(-x)dx = \langle bf, \varphi_{m,j}\rangle = \mathbb{E}(b(X_1)\varphi_{m,j}(X_1))$$

and note that, for $k \neq l$,

$$\begin{aligned}\mathbb{E}(b(X_k)u^*_{\varphi_{m,j}}(Z_k)b(X_l)\bar{u^*}_{\varphi_{m,j}}(Z_l)) &= \frac{1}{4\pi^2}\iint \mathbb{E}(b(X_k)b(X_l)e^{ixX_k - iyX_l})\varphi^*_{m,j}(-x)\varphi^*_{m,j}(y)dxdy \\ &= \mathbb{E}[b(X_k)b(X_l)\varphi_{m,j}(X_k)\varphi_{m,j}(X_l)].\end{aligned}$$

Therefore,

$$\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n b(X_k)u^*_{\varphi_{m,j}}(Z_k)\right) \leq \frac{1}{n}\mathrm{Var}(b(X_1)u^*_{\varphi_{m,j}}(Z_1)) + \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n b(X_k)\varphi_{m,j}(X_k)\right).$$

The last term requires a covariance inequality for mixing variables (Delyon (1990), Viennet (1997), Theorem 9.1 in the appendix) and uses the fact that the X_i's are β-mixing with coefficients $\beta(k)$.

$$\sum_{j\in\mathbb{Z}}\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n b(X_i)\varphi_{m,j}(X_i)\right] \leq \sum_{j\in\mathbb{Z}}\frac{4}{n}\int \beta(x)b^2(x)|\varphi_{m,j}(x)|^2 f(x)dx \leq \frac{4m}{n}\int \beta(x)b^2(x)f(x)dx$$

where β is a nonnegative function such that $\mathbb{E}(\beta^p(X)) \leq p\sum_{k\geq 0}(k+1)^{p-1}\beta(k)$ and by using that $\|\sum_j |\varphi_{m,j}|^2(.)\|_\infty = m$. Therefore if $\mathbb{E}(b^4(X_1)) < \infty$ and $\theta > 1$, then

$$\sum_{j\in\mathbb{Z}}\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n b(X_i)\varphi_{m,j}(X_i)\right] \leq \frac{4m\sqrt{2\sum_{k\geq 0}(k+1)\beta(k)\mathbb{E}(b^4(X_1))}}{n}.$$

Moreover $\sum_{j\in\mathbb{Z}}\mathrm{Var}\left(b(X_1)u^*_{\varphi_{m,j}}(Z_1)\right) \leq \mathbb{E}\left(b^2(X_1)\sum_{j\in\mathbb{Z}}(u^*_{\varphi_{m,j}}(Z_1))^2\right)$ so that

$$\sum_{j\in\mathbb{Z}}\frac{1}{n}\mathrm{Var}\left(b(X_1)u^*_{\varphi_{m,j}}(Z_1)\right) \leq \frac{\mathbb{E}(b^2(X_1))\Delta(m)}{n},$$

which gives the result. □

6.2. **Proof of Theorem 3.1.** The proof could be sketched as follows. Let us define for $m, m' \in \mathcal{M}_n$, $B_m(0,1) = \{t \in S_m, \|t\| = 1\}$ and $B_{m,m'}(0,1) = \{t \in S_m + S_{m'}, \|t\| = 1\}$. Under the definition of $\hat{m}$, $\forall m \in \mathcal{M}_n$, $\gamma_n(\hat{\ell}_{\hat{m}}) + \mathrm{pen}(\hat{m}) \leq \gamma_n(\ell_m) + \mathrm{pen}(m)$. For all functions s and t, $\gamma_n(t) - \gamma_n(s) = \|t-\ell\|^2 - \|s-\ell\|^2 - 2\nu_n(t-s)$, and

$$2\nu_n(\hat{\ell}_{\hat{m}} - \ell_m) \leq \frac{1}{4}\|\hat{\ell}_{\hat{m}} - \ell_m\|^2 + 4\sup_{t\in B_{m,\hat{m}}(0,1)}\nu_n^2(t).$$

Thus, we obtain, as $\|\hat{\ell}_{\hat{m}} - \ell_m\|^2 \leq 2\|\hat{\ell}_{\hat{m}} - \ell\|^2 + 2\|\ell_m - \ell\|^2$ that

$$\frac{1}{2}\|\hat{\ell}_{\hat{m}} - \ell\|^2 \leq \frac{3}{2}\|\ell_m - \ell\|^2 + \mathrm{pen}(m) + 4\sup_{t\in B_{m,\hat{m}}(0,1)}\nu_n^2(t) - \mathrm{pen}(\hat{m}). \tag{25}$$

Then we need to find a function $p(m,m')$ such that

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat m}(0,1)} \nu_n^2(t) - p(m,\hat m)\right)_+ \le \frac{C}{n} \tag{26}$$

which in turn will fix the penalty function through the requirement: $\forall m, m' \in \mathcal{M}_n$,

$$4p(m,m') \le \text{pen}(m) + \text{pen}(m'). \tag{27}$$

Gathering (25), (26) and (27) will lead to, $\forall m \in \mathcal{M}_n$,

$$\frac{1}{2}\mathbb{E}\left(\|\hat\ell_{\hat m} - \ell\|^2\right) \le \frac{3}{2}\|\ell_m - \ell\|^2 + 2\text{pen}(m) + \frac{4C}{n}$$

which is the result.

Now, if ν_n is split into several terms, deduced from the first decomposition given by (23)-(24), say $\nu_n(t) = \sum_{i=1}^3 \sum_{j=1}^{p_i} \nu_n^{(i,j)}(t)$ where $p_i \le 3$, then, up to some multiplicative constants, inequality (26) will follow from inequalities

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat m}(0,1)} [\nu_n^{(i,j)}(t)]^2 - p_{i,j}(m,\hat m)\right)_+ \le \frac{C_{i,j}}{n},$$

with $C = 9\sum_{i,j} C_{i,j}$ and $p(m,m') = 9\sum_{i,j} p_{i,j}(m,m')$. The study of the $\nu_n^{(i,j)}(t)$ is explained below.

First we split $\nu_n^{(1)}$ in two parts, so that both expressions involve independent variables, conditionally to (X): $\nu_n^{(1)} = \nu_n^{(1,odd)} + \nu_n^{(1,even)}$ where

$$\nu_n^{(1,even)}(t) = \frac{1}{n}\sum_{1\le 2k\le n} \varepsilon_{2k+1} u_t^*(Z_{2k}), \quad \nu_n^{(1,odd)}(t) = \frac{1}{n}\sum_{1\le 2k+1\le n} \varepsilon_{2k+2} u_t^*(Z_{2k+1}).$$

Now, we shall study $\nu_n^{(1,even)}$ only since both terms lead to the same type of result. As Talagrand's Inequality requires the random variables involved to be bounded, we have an additional step that allows to obtain the result under a moment condition on the ε_i's: $\nu_n^{(1,even)} = \nu_n^{(1,1)} + \nu_n^{(1,2)} + \nu_n^{(1,3)}$ with

$$\nu_n^{(1,1)}(t) = \frac{1}{n}\sum_{1\le 2k\le n}\left[\varepsilon_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|\le n^{1/4}} u_t^*(Z_{2k}) - \mathbb{E}_X(\varepsilon_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|\le n^{1/4}} u_t^*(Z_{2k}))\right],$$

$$\nu_n^{(1,2)}(t) = \frac{1}{n}\sum_{1\le 2k\le n} \mathbb{E}(\varepsilon_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|\le n^{1/4}})\left[t(X_{2k}) - \mathbb{E}(t(X_{2k}))\right]$$

and

$$\nu_n^{(1,3)}(t) = \frac{1}{n}\sum_{1\le 2k\le n}\left[\varepsilon_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|> n^{1/4}} u_t^*(Z_{2k}) - \mathbb{E}(\varepsilon_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|> n^{1/4}} u_t^*(Z_{2k}))\right],$$

where $\mathbb{E}_X$ denotes the conditional expectation given $(X_k)_{1\le k\le n+1}$. It is worth noticing that $\nu_n^{(1,2)}$ vanishes if the ε's are symmetric and $\nu_n^{(1,3)}(t)$ is negligible under adequate moment conditions on the ε's.

The following lemmas are proved below:

Lemma 6.2.

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}(0,1)} [\nu_n^{(1,1)}(t)]^2 - p_{1,1}(m,\hat{m})\right)_+ \leq \frac{C}{n},$$

where $p_{1,1}(m,m') = K\mathbb{E}(\varepsilon_1^2)\Psi(m \vee m')$ *where* K *is a numerical constant and* $\Psi(m)$ *is defined by (11).*

Lemma 6.3. *If the process* (X_k) *is geometrically* β*-mixing (or arithmetically with* $\theta > 3$*), then*

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}(0,1)} [\nu_n^{(1,2)}(t)]^2 - K\mathbb{E}(|\varepsilon_1|)\sum_k \beta(k)\frac{m+\hat{m}}{n}\right)_+ \leq \frac{C}{n}.$$

Lemma 6.4. *If* $\mathbb{E}(\varepsilon_1^6) < +\infty$*, and* m_n *is the largest value of* m *such that* $\Delta(m_n)/n \leq 1$*, then,*

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}(0,1)} [\nu_n^{(1,3)}(t)]^2\right) \leq \mathbb{E}\left(\sup_{t\in B_{m_n}(0,1)} (\nu_n^{(1,3)}(t))^2\right) \leq \frac{2\mathbb{E}(\varepsilon_1^6)}{n}.$$

For the study of $\nu_n^{(2)}(t)$, a result is given, whose proof is detailed in Section 8:

Lemma 6.5. *Let* $\tau_n(t) = \nu_n^{(2)}(t) = (1/n)\sum_{k=1}^n \xi_{k+1}\sigma(X_k)u_t^*(Z_k)$*. Under the assumptions of Theorem 3.1,*

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}(0,1)} [\tau_n(t)]^2 - p_\tau(m,\hat{m})\right)_+ \leq \frac{C}{n},$$

where $p_\tau(m,m') = \kappa\mathbb{E}(\sigma^2(X_1))\Psi(m\vee m')$*.*

For $\nu_n^{(3)}(t)$ we write $\nu_n^{(3)}(t) = \nu_n^{(3,1)}(t) + \nu_n^{(3,2)}(t)$ with

$$\nu_n^{(3,1)}(t) = \frac{1}{n}\sum_{k=1}^n [b(X_k)u_t^*(Z_k) - b(X_k)t(X_k))], \quad \nu_n^{(3,2)}(t) = \frac{1}{n}\sum_{k=1}^n [b(X_k)t(X_k) - \langle t,\ell\rangle],$$

where $b(X_k)t(X_k) = \mathbb{E}_{(X)}[b(X_k)u_t^*(Z_k)]$. For $\nu_n^{(3,1)}(t)$ we can apply Talagrand's Inequality conditionally to (X), for $\nu_n^{(3,2)}(t)$, we can use approximation techniques. More precisely, using the same techniques as previously, we get

Lemma 6.6. *If* $\mathbb{E}(b^8(X_1)) < +\infty$*, and* $(X_i)_{i\in\mathbb{N}}$ *is arithmetically* β*-mixing with* $\theta > 14$*, then*

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}(0,1)} [\nu_n^{(3,1)}(t)]^2 - p_{3,1}(m,\hat{m})\right)_+ \leq \frac{C}{n},$$

where $p_{3,1}(m,m') = K\mathbb{E}(b^2(X_1))\Psi(m\vee m')$*, and*

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}(0,1)} [\nu_n^{(3,2)}(t)]^2 - K'(\mathbb{E}(b^4(X_1))\sum_k (k+1)\beta(k))^{1/2}\frac{m+\hat{m}}{n}\right)_+ \leq \frac{C}{n},$$

where κ *and* κ' *are numerical constants.*

The proof of the result concerning $\nu_n^{(3,1)}$ follows the same line as the proof of Lemma 6.5, which is detailed in section 8 and is therefore omitted here. For $\nu_n^{(3,2)}$, the bound can be obtained directly by applying Talagrand's inequality (see Theorem 9.2) to this process, if b is bounded. As this is not assumed, we write $b = b_1 + b_2$ with $b_1(x) = b(x)\mathbf{1}_{|b(x)|\le n^{1/4}}$ and $b_2(x) = b(x)\mathbf{1}_{|b(x)|>n^{1/4}}$. This allows to split the process in two parts and consequently to obtain the result under $\mathbb{E}(|b(X_1)|^8) < +\infty$ and $m_n \le \sqrt{n}$, where m_n is the largest over the $m \in \mathcal{M}_n$ (a condition which is fulfilled in our problem).

Proof of Lemma 6.2.
We apply Lemma 9.2 to process $\nu_n^{(1,1)}(t)$ conditionally to the sequence $(X_k)_{1\le k\le n}$. Given the X_i's, the variables $(Z_{2k}, \varepsilon_{2k+1})_{k\ge 1}$ are independent and we have, for $m^\star = m \vee m'$,

$$\begin{aligned}
&\mathbb{E}_X\left(\sup_{t\in B_{m,m'}(0,1)} (\nu_n^{(1,1)}(t))^2\right)\\
\le\;& \sum_{j\in\mathbb{Z}} \mathbb{E}_X\left[\left(\frac{1}{n}\sum_{1\le 2k\le n} \varepsilon_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|\le n^{1/4}} u^*_{\varphi_{m^\star,j}}(Z_{2k}) - \mathbb{E}_X(\varepsilon_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|\le n^{1/4}} u^*_{\varphi_{m^\star,j}}(Z_{2k}))\right)^2\right]\\
\le\;& \sum_{j\in\mathbb{Z}} \frac{1}{n^2}\sum_{1\le 2k\le n} \mathrm{Var}_X\left(\varepsilon_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|\le n^{1/4}} u^*_{\varphi_{m^\star,j}}(Z_{2k})\right)\\
\le\;& \sum_{j\in\mathbb{Z}} \frac{1}{n^2}\sum_{1\le 2k\le n} \mathbb{E}_X\left[\left(\varepsilon_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|\le n^{1/4}} u^*_{\varphi_{m^\star,j}}(Z_{2k})\right)^2\right]\\
=\;& \frac{1}{n^2}\sum_{1\le 2k\le n} \mathbb{E}_X\left(\varepsilon^2_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|\le n^{1/4}} \sum_{j\in\mathbb{Z}} (u^*_{\varphi_{m^\star,j}}(Z_{2k}))^2\right) \le \mathbb{E}(\varepsilon_1^2)\frac{\Delta(m^\star)}{n} := H^2,
\end{aligned}$$

as $\|\sum_j |u^*_{\varphi_{m,j}}(.)|^2\|_\infty \le \Delta(m)$ by using Parseval's formula.
Next

$$\sup_{x,y} |y 1_{|y|\le n^{1/4}} u_t^*(x)| \le n^{1/4}\|u_t^*(.)\|_\infty \le n^{1/4}\sqrt{\Delta(m^*)} := M_1.$$

Lastly, following the same method as in Comte *et al.* (2006), Lemma 4, we get

$$\begin{aligned}
&\sup_{t\in B_{m,m'}(0,1)} \frac{1}{n}\sum_k \mathrm{Var}_X(\varepsilon_{2k+1}\mathbf{1}_{|\varepsilon_{2k+1}|\le n^{1/4}} u_t^*(Z_{2k}))\\
\le\;& \sup_{t\in B_{m,m'}(0,1)} \frac{1}{n}\sum_k \mathbb{E}(\varepsilon^2_{2k+1})\mathbb{E}_X[(u_t^*(Z_{2k}))^2] \le \mathbb{E}(\varepsilon_1^2)\sqrt{\Delta_2(m^\star)}/(2\pi), \text{ where}
\end{aligned}$$

$$\Delta_2(m) = m^2 \iint \left|\frac{\varphi^*(x)\varphi^*(y)}{f_\varepsilon^*(mx)f_\varepsilon^*(my)} f_\varepsilon^*(m(x-y))\right|^2 dxdy. \tag{28}$$

Then the usual bounds for Δ_2 hold, namely, $\sqrt{\Delta_2(m^\star)} \le \Delta(m^\star)$ if $\delta > 1$ and if $\delta \le 1$, $\sqrt{\Delta_2(m^\star)}/2\pi \le \kappa\Delta(m^\star)/(m^\star)^{(1-\delta)/2}$. This gives $v = c\Delta(m^*)(m^\star)^{-(1-\delta)_+/2}$.

Given that the orders are the same as in Comte *et al.* (2006) for v and H^2 and inserting the slight difference on M_1, it can easily be checked that the conclusion still holds and therefore the result of Lemma 6.2 follows. □

Proof of Lemma 6.3.
The result given in Lemma 6.3 is a standard result of density estimation for mixing variables. We refer the reader to Tribouley and Viennet (1998) or Comte and Merlevède (2002), p.217.□

Proof of Lemma 6.4. Let $e_k = \varepsilon_k \mathbf{1}_{|\varepsilon_k|>n^{1/4}}$.

$$\mathbb{E}\left(\sup_{t\in B_{m_n}(0,1)} (\nu_n^{(1,3)}(t))^2\right) \leq \sum_{j\in\mathbb{Z}} \mathrm{Var}\left(\frac{1}{n}\sum_k e_{2k+1}u^*_{\varphi_{m_n,j}}(Z_{2k})\right)$$
$$= \frac{1}{n^2}\sum_{j\in\mathbb{Z}}\left[\sum_k \mathrm{Var}\left(e_{2k+1}u^*_{\varphi_{m_n,j}}(Z_{2k})\right) + \sum_{k\neq l}\mathrm{cov}(e_{2k+1}u^*_{\varphi_{m_n,j}}(Z_{2k}), e_{2l+1}u^*_{\varphi_{m_n,j}}(Z_{2l}))\right]$$
$$\leq \frac{\mathbb{E}(e_3^2)\Delta(m_n)}{n} + \frac{1}{n^2}\sum_{j\in\mathbb{Z}}\mathbb{E}(e_3)^2\sum_{k\neq l}\mathrm{cov}(\varphi_{m_n,j}(X_{2k}),\varphi_{m_n,j}(X_{2l}))$$
$$\leq \frac{\mathbb{E}(e_3^2)\Delta(m_n)}{n} + \frac{1}{n^2}\sum_{j\in\mathbb{Z}}\mathbb{E}(e_3)^2\mathrm{Var}\left(\sum_k \varphi_{m_n,j}(X_{2k})\right)$$
$$\leq \mathbb{E}(e_3^2)\left(\frac{\Delta(m_n)}{n} + \frac{m_n\sum_k\beta_k}{n}\right) \leq \frac{2\Delta(m_n)}{n}\mathbb{E}\left(\varepsilon_1^2\mathbf{1}_{|\varepsilon_1|\geq n^{1/4}}\right) \leq 2\mathbb{E}\left(\varepsilon_1^6/n\right).\Box$$

7. Proof of Proposition 4.1.

We bound the expectations of the empirical processes involved in order to obtain the bound of $\sum_{j\in\mathbb{Z}}\mathbb{E}(\breve{\nu}_n(\varphi_{m,j}))$, using the decomposition $\breve{\nu}_n = \sum_{i=1}^4 \breve{\nu}_n^{(i)}$ with

$$\breve{\nu}_n^{(1)}(t) = (1/n)\sum_{k=1}^n\left[(b^2(X_k)+\sigma^2(X_k))u_t^*(Z_k) - \langle t,\vartheta\rangle\right],$$

$\breve{\nu}_n^{(2)}(t) = (1/n)\sum_{k=1}^n(\xi_{k+1}^2-1)\sigma^2(X_k)u_t^*(Z_k)$, $\breve{\nu}_n^{(3)}(t) = \frac{1}{n}\sum_{k=1}^n(\varepsilon_{k+1}^2 - s_\varepsilon^2)u_t^*(Z_k)$, and $\breve{\nu}_n^{(4)}(t) = (2/n)\sum_{k=1}^n[\varepsilon_{k+1}\xi_{k+1}\sigma(X_k)+b(X_k)\varepsilon_{k+1}+\sigma(X_k)b(X_k)\xi_{k+1}]u_t^*(Z_k)$. But it is clear that $\breve{\nu}_n^{(1)}$ is the same process as $\nu_n^{(3)}$ with $b(X_k)$ replaced by $(b^2+\sigma^2)(X_k)$, that $\breve{\nu}_n^{(2)}$ is of the same type as $\nu_n^{(2)}$ with $\sigma(X_k)$ replaced by $\sigma^2(X_k)$ and ξ_{k+1} by ξ_{k+1}^2-1. Next, $\breve{\nu}_n^{(3)}$ corresponds to $\nu_n^{(1)}$ with ε_{k+1} replaced by $\varepsilon_{k+1}^2 - s_\varepsilon^2$. Lastly

$$\sum_{j\in\mathbb{Z}}\mathbb{E}[(\breve{\nu}_n^{(4)}(\varphi_{m,j}))^2] = \frac{4}{n}\sum_{j\in\mathbb{Z}}\mathbb{E}[(\varepsilon_2\xi_2\sigma(X_1)+b(X_1)\varepsilon_2+(\sigma b)(X_1)\xi_2)^2(u^*_{\varphi_{m,j}}(Z_k))^2]$$
$$\leq \frac{4\Delta(m)}{n}[s_\varepsilon^2\mathbb{E}(\sigma^2(X_1))+\mathbb{E}(b^2(X_1)(s_\varepsilon^2+\sigma^2(X_1)))].$$

The last step is to gather the terms.□

8. Proof of Lemma 6.5

If $t = t_1 + t_2$ with t_1 in S_m and t_2 in $S_{m'}$, then t is such that t^* has its support included in $[-\pi \max(m, m'), \pi \max(m, m')]$ and therefore t belongs to $S_{m^\star}$ where $m^\star = \max(m, m')$. We recall that $B_{m,m'}(0, 1) = \{t \in S_{m^\star} \ / \ \|t\| = 1\}$. Denote by

$$\mathbb{H}_\tau^2(m, m') \quad = \quad (n^{-1} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i))\Delta(m^\star)/n, \tag{29}$$

and let $\sigma_\tau^2 = \mathbb{E}(\xi_2^2)\mathbb{E}(\sigma^2(X_1)) = \mathbb{E}(\sigma^2(X_1))$. We have

$$\mathbb{H}_\tau^2(m, m') = (n^{-1} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2)\Delta(m^\star)/n + \sigma_\tau^2 \Delta(m^\star)/n,$$

which is bounded by $\mathbb{H}_{\tau,1}(m, m') + \mathbb{H}_{\tau,2}(m, m')$ where

$$\mathbb{H}_{\tau,1}(m, m') = (n^{-1} \sum_{i=1}^n \xi_{i+1}^2 \sigma^2(X_i) - \sigma_\tau^2)\mathbf{1}_{\{|n^{-1}\sum_{i=1}^n \xi_{i+1}^2\sigma^2(X_i) - \sigma_\tau^2| > \sigma_\tau^2/2\}} \frac{\Delta(m^\star)}{n}$$

and $\mathbb{H}_{\tau,2}(m, m') = 3\sigma_\tau^2 \Delta(m^\star)/(2n)$. We infer that $\tau_n(t) = \tau_n^{(1)}(t) + \tau_n^{(2)}(t)$ with

$$\tau_n^{(1)}(t) = \frac{1}{n}\sum_{k=1}^n \xi_{k+1}\sigma(X_k)\left[u_t^*(Z_k) - t(X_k)\right], \ \tau_n^{(2)}(t) = \frac{1}{n}\sum_{k=1}^n \xi_{k+1}\sigma(X_k)t(X_k),$$

and

$$\begin{aligned}
& \mathbb{E}[\sup_{t \in B_{m,m'}(0,1)} |\tau_n(t)|^2 - p_\tau(m, m')]_+ \\
\leq \ & 2\mathbb{E}[\sup_{t \in B_{m,m'}(0,1)} |\tau_n^{(1)}(t)|^2 - 2(1 + 2\epsilon^2(m, m'))\mathbb{H}_\tau^2(m, m')]_+ \\
& +2\mathbb{E}[\sup_{t \in B_{m,m'}(0,1)} |\tau_n^{(2)}(t)|^2 - 6p_2(m, m')]_+ \\
& +\mathbb{E}[4(1 + 2\epsilon^2(m, m'))\mathbb{H}_\tau^2(m, m') + 12p_2(m, m') - p_\tau(m, m')]_+,
\end{aligned}$$

where $\epsilon(m, m') = 1$ if $0 \leq \delta < 1/3$ and $\epsilon(m, m') = C(f_\varepsilon)(\pi(m \vee m'))^{1/2-(1/2-\delta/2)_+}$ (see Comte *et al.* (2006)) and

$$p_2(m, m') = \mathbb{E}(\sigma^2(X_1))\frac{m^\star}{n}. \tag{30}$$

Clearly, $p_2(m, m')$ is negligible with respect to $p_\tau(m, m')$ (compare their orders in $m^\star$), so that for simplicity we consider that $(12p_2(m, m') - p_\tau(m, m')/2)_+ \leq C/n$.

$$\begin{aligned}
&\mathbb{E}[4(1 + 2\epsilon^2(m, m'))\mathbb{H}_\tau^2(m, m') + 12p_2(m, m') - p_\tau(m, m')]_+ \\
&\leq 4(1 + 2\epsilon^2(m, m'))\mathbb{E}|\mathbb{H}_{\tau,1}(m, m')| + \mathbb{E}[4(1 + 2\epsilon^2(m, m'))\mathbb{H}_{\tau,2}(m, m') - p_\tau(m, m')/2]_+.
\end{aligned} \tag{31}$$

Since we only consider values of m such that the penalty are bounded by some constant K, we obtain that for some $p \geq 2$, $\mathbb{E}|\mathbb{H}_{\tau,1}(m,m')|$ is bounded by

$$\begin{aligned} & C\mathbb{E}\left[|\frac{1}{n}\sum_{i=1}^{n}\xi_{i+1}^2\sigma^2(X_i) - \sigma_\tau^2|\mathbf{I}_{\{n^{-1}|\sum_{i=1}^n(\xi_{i+1}^2\sigma^2(X_i)-\sigma_\tau^2)|>\sigma_\tau^2/2\}}\right] \\ \leq \quad & C2^{p-1}\mathbb{E}\left[|n^{-1}\sum_{i=1}^{n}\xi_{i+1}^2\sigma^2(X_i) - \sigma_\tau^2|^p\right]/\sigma_\tau^{2(p-1)}. \end{aligned}$$

Moreover, we shall see below that $\epsilon(m,m')$ is constant (if $\delta = 0$ or $0 < \delta < 1/3$) or at most of order $(\ln(n))^\delta$ (if $\delta > 1/3$). According to Rosenthal's inequality (see Rosenthal (1970)) generalized to the mixing case (see Doukhan (1994) and Inequality (34) recalled in Lemma 9.1), we find that,

$$\mathbb{E}|n^{-1}\sum_{i=1}^{n}\xi_{i+1}^2\sigma^2(X_i) - \sigma_\tau^2|^p \leq C'(p,\xi,\sigma(X))\left(n^{1-p} + n^{-p/2}\right).$$

Now, Assumption **A1**(i)-**A5** implies that $\gamma > 1/2$, therefore $|\mathcal{M}_n| \leq \sqrt{n}$ if $\delta = 0$ and has logarithmic order if $\delta > 0$ and thus, choosing $p = 3$ leads to $\sum_{m'\in\mathcal{M}_n}\mathbb{E}|(1+2\epsilon^2(m,m'))\mathbb{H}_{\tau,1}(m,m')| \leq C(\xi,\sigma(X))/n$, where $C(\xi,\sigma(X))$ is a constant depending on the moments of ξ_1 and $\sigma(X_1)$. In particular this requires that ξ admit a moment of order 8.

The last term of the inequality (31) vanishes as soon as

$$p_\tau(m,m') = 8(1+2\epsilon^2(m,m'))\mathbb{H}_{\tau,2}(m,m') = 12(1+2\epsilon^2(m,m'))\mathbb{E}(\sigma^2(X_1))\Delta(m^\star)/n.$$

For this choice of $p_\tau(m,m')$, we obtain that

$$\begin{aligned} & \mathbb{E}[\sup_{t\in B_{m,\hat{m}}(0,1)}|\tau_n(t)|^2 - p_\tau(m,\hat{m})]_+ \\ \leq \quad & 2\sum_{m'\in\mathcal{M}_n}\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}(n^{-1}\sum_{i=1}^{n}\xi_i\sigma(X_i)(u_t^*(Z_i)-t(X_i)))^2 - 2(1+2\epsilon^2(m,m'))\mathbb{H}_\tau^2(m,m')]_+ \\ & +2\sum_{m'\in\mathcal{M}_n}\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}|\tau_n^{(2)}(t)|^2 - 6p_2(m,m')]_+ + \frac{C}{n}. \end{aligned}$$

Then we apply the following Lemma.

Lemma 8.1. *Under the assumptions on the model, if $\mathbb{E}|\xi_1|^8 < \infty$ and $\mathbb{E}(\sigma^8(X_1))$, then for some given $\epsilon > 0$:*

$$\begin{aligned} & \sum_{m'\in\mathcal{M}_n}\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_{i+1}\sigma(X_i)(u_t^*(Z_i)-t(X_i))\right)^2 - 2(1+2\epsilon^2)\mathbb{H}_\tau^2(m,m')\right]_+ \\ & \qquad \leq K_1\left\{\sum_{m'\in\mathcal{M}_n}\left[\frac{\sigma_\tau^2\lambda_2\Gamma_2(m^\star)}{n}\exp\left(-K_2\epsilon^2\frac{\Delta(m^\star)}{\lambda_2\Gamma_2(m^\star)}\right)\right] + \left(1+\frac{\ln^4(n)}{\sqrt{n}}\right)\frac{1}{n}\right\}, \end{aligned}$$

where λ_2 is a constant, $\Gamma_2(m)$ is defined by

$$\Gamma_2(m) = (m)^{2\gamma+\min[(1/2-\delta/2),(1-\delta)]}\exp\{2\mu(\pi m)^\delta\} \tag{32}$$

and K_1 and K_2 are constants depending on the moments of ξ and $\sigma(X)$.

Moreover, it also follows from Baraud *et al.* (2001) and Comte and Rozenholc (2002), that the process $\tau_n^{(2)}$ is a standard process of the auto-regressive context and satisfies, for $p_2(m,m')$ defined by (30),

$$2\sum_{m'\in\mathcal{M}_n}\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}|\tau_n^{(2)}(t)|^2-6p_2(m,m')]_+\leq\frac{c}{n}.$$

We denote by

$$A(m)=\frac{K_1\sigma_\tau^2}{n}\lambda_2\Gamma_2(m)\exp\left(-K_2\epsilon^2\frac{\Delta(m)}{\lambda_2\Gamma_2(m)}\right)=\frac{K_1\sigma_\tau^2\lambda_2\Gamma_2(m)}{n}\exp\left(-\kappa_2\epsilon^2m^{(1/2-\delta/2)_+}\right).$$

The study of $A(m^\star)$ is standard in deconvolution (see Comte *et al.* (2006)) and leads to choose $\epsilon^2(m,m')$ as a constant if $\delta\leq 1/3$ and of order $m^{\delta-(1/2-\delta/2)_+}$ if $\delta>1/3$, to ensure that $\sum_{m'\in\mathcal{M}_n}A(m^\star)$ is less than C/n.

With $p_\tau(m,m')$ given in Lemma 6.5, by gathering all terms we find the result. $\square$

Proof of Lemma 8.1.

We work conditionally to the (ξ_i,X_i)'s and $\mathbb{E}_{X,\xi}$ and $\mathbb{P}_{X,\xi}$ denote the conditional expectations and probability for fixed $\xi_1,\dots,\xi_n,\xi_{n+1},X_1,\dots,X_n$.

We apply Lemma 9.2 with $f_t(\xi_i,X_i,Z_i)=\xi_{i+1}\sigma(X_i)u_t^*(Z_i)$, conditionally to the ξ_i's and X_i's to the random variables $(\xi_2,X_1,Z_1),\dots,(\xi_{n+1},X_n,Z_n)$ which are independent but non identically distributed since the ξ_i's and the $X_i's$ are fixed constants.

Straightforward calculations give that for $\mathbb{H}_\tau(m,m')$ defined in (29) we have

$$\mathbb{E}^2_{X,\xi}[\sup_{t\in B_{m,m'}(0,1)}n^{-1}\sum_{l=1}^n\xi_{l+1}\sigma(X_l)(u_t^*(Z_l)-t(X_l))]\leq\mathbb{H}_\tau^2(m,m').$$

Let $P_{j,k}^{(l)}(m)=\mathbb{E}_{X,\xi}[u^*_{\varphi_{m,j}}(Z_l)u^*_{\varphi_{m,k}}(-Z_l)]$. Write

$$\sup_{t\in B_{m,m'}(0,1)}\frac{1}{n}\sum_{l=1}^n\mathrm{Var}_{X,\xi}(\xi_{l+1}\sigma(X_l)u_t^*(Z_l))\quad\leq\quad\frac{1}{n}\sum_{l=1}^n\xi_{l+1}^2\sigma^2(X_l)\left(\sum_{j,k\in\mathbb{Z}}|P_{j,k}^{(l)}(m^\star)|^2\right)^{1/2}.$$

We argue as in Comte *et al.* (2006). Let recall that $\Delta_2(m)$ is defined by (28). We have $\Delta_2(m)\leq\lambda_2^2\Gamma_2^2(m)$, with Γ_2 defined by (32) and $\lambda_2=\lambda_2(\gamma,A_0,\delta,\mu,\|f_\varepsilon\|)$. Now, write $P_{j,k}^{(l)}$ as

$$P_{j,k}^{(l)}(m)\quad=\quad\frac{m}{4\pi^2}\iint\frac{e^{-ixj-iyk}e^{im(x-y)X_l}\varphi^*(-x)\varphi^*(-y)}{f_\varepsilon^*(mx)f_\varepsilon^*(my)}f_\varepsilon^*(m(x-y))dxdy.$$

By applying Parseval's formula we get that $\sum_{j,k}|P_{j,k}^{(l)}(m)|^2$ equals $\Delta_2(m)$. We now write that
$\sup_{t\in B_{m,m'}(0,1)}(n^{-1}\sum_{i=1}^n\mathrm{Var}_{X,\xi}(\xi_{i+1}\sigma(X_i)u_t^*(Z_i)))\leq(n^{-1}\sum_{i=1}^n\xi_{i+1}^2\sigma^2(X_i))\lambda_2\Gamma_2(m^\star)$, and thus we take $v_\tau(m,m')=(n^{-1}\sum_{i=1}^n\xi_{i+1}^2\sigma^2(X_i))\lambda_2\Gamma_2(m^\star)$. Lastly, since $\sup_{t\in B_{m,m'}(0,1)}\|f_t\|_\infty\leq 2\max_{1\leq i\leq n}|\xi_{i+1}\sigma(X_i)|\sqrt{\Delta(m^\star)}$, we take $M_{1,\tau}(m,m')=2\max_{1\leq i\leq n}|\xi_{i+1}\sigma(X_i)|\sqrt{\Delta(m^\star)}$.

By applying Lemma 9.2, we get for some constants κ_1, κ_2, κ_3

$$\begin{aligned}
&\mathbb{E}_{X,\xi}[\sup_{t\in B_{m,m'}(0,1)} \nu_{n,1}^2(t) - 2(1+2\epsilon^2)\mathbb{H}_\tau^2]_+ \\
&\qquad \le K_1 \left[\frac{\lambda_2\Gamma_2(m^\star)}{n^2}(\sum_{i=1}^n \xi_{i+1}^2\sigma^2(X_i)) \exp\left\{-K_2\epsilon^2\frac{\Delta(m^\star)}{\lambda_2\Gamma_2(m^*)}\right\}\right. \\
&\qquad\qquad \left. + \frac{\Delta(m^\star)}{n^2}(\max_{1\le i\le n} \xi_{i+1}^2\sigma^2(X_i)) \exp\left\{-K_3\epsilon C(\epsilon^2)\frac{\sqrt{\sum_{i=1}^n \xi_{i+1}^2\sigma^2(X_i)}}{\max_i |\xi_{i+1}\sigma(X_i)|}\right\}\right]
\end{aligned}$$

To relax the conditioning, it suffices to integrate with respect to the law of the (ξ_{i+1}, X_i)'s the above expression. The first term in the bound simply becomes:

$$\sigma_\tau^2\lambda_2\Gamma_2(m^\star)\exp[-\kappa_2\epsilon\Delta(m^\star)/(\lambda_2\Gamma_2(m^\star))]/n.$$

The second term is bounded by

$$\frac{\Delta(m^\star)}{n^2}\mathbb{E}\left[(\max|\xi_{i+1}\sigma(X_i)|^2)\exp\left(-\kappa_3\epsilon C(\epsilon^2)\frac{\sqrt{\sum_{i=1}^n \xi_{i+1}^2\sigma^2(X_i)}}{\max_{1\le i\le n}|\xi_{i+1}\sigma(X_i)|}\right)\right]. \tag{33}$$

Since we only consider integers m such that the penalty term is bounded, we have $\Delta(m)/n \le K$ and the sum of the above terms for $m' \in \mathcal{M}_n$ and $|\mathcal{M}_n| \le \sqrt{n}$ is less than

$$\frac{K}{\sqrt{n}}\mathbb{E}\left[\left(\max_{1\le i\le n}\xi_{i+1}^2\sigma^2(X_i)\right)\exp\left(-\kappa_3\epsilon C(\epsilon^2)\frac{\sqrt{\sum_{i=1}^n \xi_{i+1}^2\sigma^2(X_i)}}{\max_{1\le i\le n}|\xi_{i+1}\sigma(X_i)|}\right)\right].$$

We need to study when such a term is less than c/n for some constant c. We bound $\max_i|\xi_{i+1}\sigma(X_i)|$ by $m_{\xi,\sigma}$ on the set $\{\max_i|\xi_{i+1}\sigma(X_i)| \le m_{\xi,\sigma}\}$ and the exponential by 1 on the set $\{\max_i|\xi_{i+1}\sigma(X_i)| > m_{\xi,\sigma}\}$ and by denoting $\mu_\epsilon = \kappa_3\epsilon C(\epsilon^2)$, this yields

$$\begin{aligned}
&\mathbb{E}\left[\max_{1\le i\le n}\xi_{i+1}^2\sigma^2(X_i)\exp\left(-\mu_\epsilon\sqrt{\frac{\sum_{i=1}^n \xi_{i+1}^2\sigma^2(X_i)}{\max_{1\le i\le n}\xi_{i+1}^2\sigma^2(X_i)}}\right)\right] \\
&\le m_{\xi,\sigma}^2\mathbb{E}\left(\exp(-\mu_\epsilon\frac{\sqrt{\sum_{i=1}^n \xi_{i+1}^2\sigma^2(X_i)}}{m_{\xi,\sigma}})\right) + \mathbb{E}\left(\max_{1\le i\le n}\xi_{i+1}^2\sigma^2(X_i)\mathbf{I}_{\{\max_{1\le i\le n}|\xi_{i+1}\sigma(X_i)|>m_{\xi,\sigma}\}}\right) \\
&\le \quad m_{\xi,\sigma}^2\left[\mathbb{E}\left(\exp(-\mu_\epsilon\sqrt{n\sigma_\tau^2/(2m_{\xi,\sigma}^2)})\right) + \mathbb{P}\left(|\frac{1}{n}\sum_{i=1}^n\xi_{i+1}^2\sigma^2(X_i) - \sigma_\tau^2| \ge \sigma_\tau^2/2\right)\right] \\
&\qquad + m_{\xi,\sigma}^{-r}\mathbb{E}(\max_{1\le i\le n}|\xi_{i+1}\sigma(X_i)|^{r+2}) \\
&\le \quad m_{\xi,\sigma}^2 e^{-\mu_\epsilon\sqrt{n}\sigma_\tau/(\sqrt{2}m_{\xi,\sigma})} + m_{\xi,\sigma}^2 2^p\sigma_\tau^{-2p}\mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^n\xi_{i+1}^2\sigma^2(X_i) - \sigma_\tau^2\right|^p\right) \\
&\qquad + m_{\xi,\sigma}^{-r}\mathbb{E}(\max_{1\le i\le n}|\xi_{i+1}\sigma(X_i)|^{r+2}).
\end{aligned}$$

Again by applying Rosenthal's inequality (see Lemma 9.1), we obtain that

$$\mathbb{E}\left[\max_{1\le i\le n}\xi_{i+1}^2\sigma^2(X_i)\exp\left(-\mu_\epsilon\sqrt{\frac{\sum_{i=1}^n\xi_{i+1}^2\sigma^2(X_i)}{\max_{1\le i\le n}\xi_{i+1}^2\sigma^2(X_i)}}\right)\right]$$
$$\le m_{\xi,\sigma}^2 e^{-\mu_\epsilon\sqrt{n}\sigma_\tau/(\sqrt{2}m_{\xi,\sigma})}+m_{\xi,\sigma}^2\frac{C(p,\xi,\sigma(X))}{n^p}[n+n^{p/2}]+n\mathbb{E}(|\xi_2\sigma(X_1)|^{r+2})m_{\xi,\sigma}^{-r}$$

also bounded by

$m_{\xi,\sigma}^2 e^{-\mu_\epsilon\sqrt{n}\sigma_\tau/(\sqrt{2}m_{\xi,\sigma})}+C'(p,\xi,\sigma(X))m_{\xi,\sigma}^2[n^{1-p}+n^{-p/2}]+n\mathbb{E}(|\xi_1|^{r+2})\mathbb{E}(|\sigma(X_1)|^{r+2})m_{\xi,\sigma}^{-r}$.

Since $\mathbb{E}|\xi_1|^8<\infty$, we take $p=3$, $c=4$ in Lemma 9.1, $r=4$, $m_{\xi,\sigma}=\sigma_\tau\epsilon C(\epsilon^2)\kappa_3\sqrt{n}/[2\sqrt{2}\ln(n)]$ and for any $n\ge 3$, and for C_1 and C_2 some constants depending on the moments of ξ and $\sigma(X)$, we find that

$$\mathbb{E}\left\{\left(\max_{1\le i\le n}\xi_{i+1}^2\sigma^2(X_i)\right)\exp\left(-\kappa_3\epsilon C(\epsilon^2)\sqrt{\sum_{i=1}^n\frac{\xi_{i+1}^2\sigma^2(X_i)}{\max_{1\le i\le n}\xi_{i+1}^2\sigma^2(X_i)}}\right)\right\}$$
$$\le \frac{C_1}{\sqrt{n}}+C_2\left(\frac{\ln^4(n)}{\sqrt{n}}\right)\frac{1}{\sqrt{n}}.$$

Then the sum over $\mathcal{M}_n$ with cardinality less than $\sqrt{n}$ of the terms in (33) is bounded by $C(1+\ln(n)^4/\sqrt{n})/n$ for some constant C, by using again that $\Delta(m^\star)/n$ is bounded. □

9. Appendix

As a reminder, some definitions and properties related to β-mixing sequences are given in this section. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space. Let Y be a random variable with values in a Banach space $(\mathbb{B},\|\cdot\|_{\mathbb{B}})$, and let $\mathcal{M}$ be a σ-algebra of $\mathcal{A}$. Let $\mathbb{P}_{Y|\mathcal{M}}$ be a conditional distribution of Y given $\mathcal{M}$, and let P_Y be the distribution of Y. Let $\mathcal{B}(\mathbb{B})$ be the borel σ-algebra on $(\mathbb{B},\|\cdot\|_{\mathbb{B}})$. Define now $\beta(\mathcal{M},\sigma(Y))=\mathbb{E}\Big(\sup_{A\in\mathcal{B}(\mathcal{X})}|\mathbb{P}_{Y|\mathcal{M}}(A)-\mathbb{P}_Y(A)|\Big)$ The coefficient $\beta(\mathcal{M},\sigma(Y))$ is the usual mixing coefficient, introduced by Volkonskiĭ and Rozanov (1960). Let $\mathbf{X}=(X_i)_{i\ge1}$ be a strictly stationary sequence of real-valued random variables. For any $k\ge0$, the coefficients $\beta_{\mathbf{X},1}(k)$ are defined by $\beta_{\mathbf{X},1}(k)=\beta(\sigma(X_1),\sigma(X_{1+k}))$, Let $\mathcal{M}_i=\sigma(X_k,1\le k\le i)$. The coefficients $\beta_{\mathbf{X},\infty}(k)$ are defined by $\beta_{\mathbf{X},\infty}(k)=\sup_{i\ge1,l\ge1}\sup\{\beta(\mathcal{M}_i,\sigma(X_{i_1},\dots,X_{i_l})),i+k\le i_1<\dots<i_l\}$,

In the paper, we do not distinguish between the two types of mixing and denote the coefficients of the process X by $\beta(k)$ or $\beta_X(k)$. It is implicit that when only covariance inequality are involved, then the milder mixing $\beta_{X,1}(k)$ is required, and we shall assume that stronger $\beta_{X,\infty}(k)$ mixing coefficients are used in the general case.

Now, a Rosenthal-type inequality for mixing variables can be deduced from Doukhan (1994), Theorem 2 p.26 and the following result holds:

Lemma 9.1. *Let $(Y_k)_{1\le k\le n}$ be a sequence of centered and stationary β-mixing variables with coefficients $\beta(k)$, admitting moments of order $r+1$ and $r\ge2$, then if*

$$\exists c\in 2\mathbb{N}, c\ge r,\ \text{such that}\sum_{k\ge1}(k+1)^{c-2}\beta(k)^{\frac{1}{c+1}}<+\infty,$$

we have the bound

$$\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^{n} Y_k\right|^r \leq C(r)\left\{n^{1-r}[\mathbb{E}|Y_1|^{r+1}]^{r/(r+1)} + n^{-r/2}[(\mathbb{E}|Y_1|^3)]^{r/3}\right\}. \tag{34}$$

We also use Delyon's (1990) covariance Inequality, successfully exploited by Viennet (1997) for partial sums of strictly stationary processes.

Theorem 9.1. *(Delyon (1990), Viennet (1997)) Let P be the distribution of Z_0 on a probability space $\mathcal{X}$, $\int f dP = \mathbb{E}_P(f)$ for any function f P-integrable. For $r \geq 2$, let $\mathcal{L}(r, \beta, P)$ be the set of functions $b_Z : \mathcal{X} \to \mathbb{R}^+$ such that*

$$b_Z = \sum_{l\geq 0}(l+1)^{r-2} b_{l,Z} \text{ with } 0 \leq b_{l,Z} \leq 1 \text{ and } \mathbb{E}_P(b_{l,Z}) \leq \beta_Z(l)$$

We define B_r as $B_r = \sum_{l\geq 0}(l+1)^{r-2}\beta_Z(l)$. Then for $1 \leq p < \infty$ and any function b_Z in $\mathcal{L}(2, \beta, P)$, $\mathbb{E}_P(b_Z^p) \leq pB_{p+1}$, as soon as $B_{p+1} < \infty$. The following result holds for a strictly stationary absolutely regular sequence, $(Z_i)_{i\in\mathbb{Z}}$, with β-mixing coefficients $(\beta_Z(k))_{k\geq 0}$: if $B_2 < +\infty$, there exists $b_Z \in \mathcal{L}(2, \beta, \infty)$ such that for any positive integer n and any measurable function $f \in \mathbb{L}_2(P)$, we have

$$\mathrm{Var}\left(\sum_{i=1}^{n} f(Z_i)\right) \leq 4n\mathbb{E}_P(b_Z f^2) = 4n\int b_Z(x) f^2(x) dP(x).$$

Lastly, we recall the version of the Talagrand inequality that is required in the paper. Mention must be made that it is valid for independent but non necessarily identically distributed random variables, which is useful here when we work conditionally to one or two of the sequences.

Lemma 9.2. *Let $Y_1, \ldots, Y_n$ be independent random variables, let $\nu_{n,Y}(f) = (1/n)\sum_{i=1}^{n}[f(Y_i) - \mathbb{E}(f(Y_i))]$ and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\xi^2 > 0$*

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_{n,Y}(f)|^2 - 2(1+2\epsilon^2)H^2\Big]_+ \leq \frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\epsilon^2\frac{nH^2}{v}} + \frac{98M_1^2}{K_1 n^2 C^2(\epsilon^2)}e^{-\frac{2K_1C(\epsilon^2)\epsilon}{7\sqrt{2}}\frac{nH}{M_1}}\right),$$

with $C(\epsilon^2) = \sqrt{1+\epsilon^2} - 1$, $K_1 = 1/6$, and

$$\sup_{f\in\mathcal{F}}\|f\|_\infty \leq M_1, \quad \mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_{n,Y}(f)|\Big] \leq H, \ \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{k=1}^{n}\mathrm{Var}(f(Y_k)) \leq v.$$

This result follows from the concentration inequality given in Klein and Rio (2005) and arguments in Birgé and Massart (1998) (see the proof of their Corollary 2 page 354).

References

[1] Baraud, Y., Comte, F. and Viennet, G. (2001a). Adaptive estimation in an autoregressive and a geometrical β-mixing regression framework. *Annals of Statistics*, **39**, 839-875.

[2] Bennett, J., Wakefield, J. (2001) Errors-in-variables in joint population pharmacokinetic/pharmacodynamic modeling. *Biometrics* **57**, 803–812.

[3] Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.

[4] Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli* **11**, 309-340.

[5] Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.* **32**, 181-192.

[6] Butucea, C. and Tsybakov, A.B. (2008). Sharp optimality for density deconvolution with dominating bias. I. *Theor. Probab. Appl.* **52**, 24–39.

[7] Cappé, O., Moulines, E. and Ryden, T. (2005). *Inference in hidden Markov models.* Springer series en Statistics. Springer, New-York.

[8] Carrasco, M., Florens, J.-P. and Renault, E. (2007). Linear Inverse Problems in Structural Econometrics, in *Handbook of Econometrics*, volume 6, eds Heckman, J. and Leamer, E., Elsevier.

[9] Chanda, K. C. (1995). Large sample analysis of autoregressive moving average models with errors-in-variables. *J. Time Ser. Anal.*, **16**, 1, 1-15.

[10] Comte, F., Dedecker, J. and Taupin, M.-L. (2008) Adaptive density deconvolution with dependent inputs. *Math. Methods Statist.* **17**, 87-112.

[11] Comte, F. and Merlevède, F. (2002). Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.* **6**, 211-238.

[12] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2009). Nonparametric estimation for a stochastic volatility model. Working paper MAP5 2007-19. To appear in *Finance and Stochastics.*

[13] Comte, F. and Rozenholc, Y. (2002). Adaptive estimation of mean and volatility functions in (auto-) regressive models. *Stochastic Process. Appl.* **97**, 111-145.

[14] Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for density deconvolution. *Canad. J. Statist.* **34**, 431-452.

[15] Comte, F., Rozenholc, Y. and Taupin, M.-L. (2007). Finite sample penalization in adaptive density deconvolution. *J. Stat. Comput. Simul.* **7**, 977-1000.

[16] Comte, F. and Taupin, M.-L. (2001). Semi-parametric estimation in the nonlinear errors-in-variables autoregressive model. *Mathematical Methods of Statistics* **10**, 121-160.

[17] Comte, F. and Taupin, M.-L. (2007). Adaptive estimation in a nonparametric regression model with errors-in-variables. *Statistica Sinica*, **17**, 1065-1090.

[18] Delyon, B. (1990). Limit theorem for mixing processes. *Technical report IRISA Rennes*, **1**, 546.

[19] van Es, B., Spreij, P. and van Zanten, H. (2005). Nonparametric volatility density estimation for discrete time models. *J. Nonparametr. Stat.* **17**, 237-251.

[20] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.

[21] Fan, J. and Masry, M. (1992). Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *J. Multivariate Anal.* **43**, 237-272.

[22] Fan, J. and Truong, Y.K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21**, 4, 1900-1925.

[23] Fan, J., Truong, Y. K. and Wang, Y. (1991). Nonparametric function estimation involving errors-in-variables. *Nonparametric Functional Estimation and Related Topics* 613-627.

[24] Gloter, A. (2007). Efficient estimation of drift parameters in stochastic volatility models. *Finance Stoch.* **11**, 495–519.

[25] Hansen, L. P. and Scheinkman, J. A. (1995) Back to the future: generating moment implications for continuous-time Markov processes. *Econometrica* **63**, 767–804.

[26] Hansen, L. P., Scheinkman, J. A. and Touzi, N. (1998). Spectral methods for identifying scalar diffusions. *J. Econometrics* **86**, 1–32.

[27] Hong, H. and Tamer, E. (2003) A simple estimator for nonlinear error in variable models. *J. Econometrics* **117**, 1–19.

[28] Ioannides, D.A. and Alevizos, P. D. (1997). Nonparametric regression with errors in variables and applications. *Statist. Probab. Lett.* **32**, 35-43.

[29] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060-1077.

[30] Koo, J.-Y. and Lee K.-W. (1998). B-spline estimation of regression functions with errors in variable. *Statist. Probab. Lett.* **40**, 57-66.

[31] Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris* **342**, 877-882.

[32] Lacour, C. (2008). Nonparametric estimation of the stationary density and the transition of a Markov chain. *Stochastic Process. Appl.* **118**, 232-260.

[33] Meyer, Y. (1990) Ondelettes et opérateurs. I. (French) [Wavelets and operators. I] Hermann, Paris.

[34] Pensky, M. and Vidakovic, B (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27 (6)** 2033-2053.

[35] Renò, R. (2006) Nonparametric estimation of stochastic volatility models. *Econom. Lett.* **90**, 390–395.

[36] Schennach, S. M. (2007). Instrumental variable estimation of nonlinear errors-in-variables models. *Econometrica* **75**, 201–239.

[37] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In: D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (Eds). *Time Series Models in Econometrics, Finance and Other Fields* London: Chapman & Hall, 1-67.

[38] Shephard, N. (2005) *Stochastic volatility.* Selected readings. Advanced Texts in Econometrics. Oxford: Oxford University Press.

[39] Taupin, M.L. (2001). Semi-parametric estimation in nonlinear structural errors-in-variables model. *Ann. Statist.* **29 (1)** 66-93.

[40] Tribouley, K. and Viennet, G. (1998). $\mathbf{L}_p$ adaptive density estimation in a β mixing framework. *Ann. Inst. H. Poincar Probab. Statist.* **34**, 179-208.

[41] Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Relat. Fields* **107**, 467-492.

[42] Volkonskiĭ, V. and Rozanov, Y. (1960). Some limit theorems for random functions. I. *Theor. Probab. Appl.* **4**, 178-197.